HELLENIC REPUBLIC
National and Kapodistrian University of Athens
EST. 1837

National and Kapodistrian University of Athens
School of Health Sciences, School of Medicine
Sector of Social Medicine, Psychiatry, Neurology
Department of Hygiene, Epidemiology and Medical Statistics
Director: Pagona Lagiou

Statistical modeling of CD4 cell count evolution in HIV-positive individuals before and after antiretroviral treatment initiation under missing data due to various dropout mechanisms

(Μοντέλα διαχρονικής εξέλιξης CD4 λεμφοκυττάρων σε HIV οροθετικά άτομα πριν και μετά την έναρξη αντιρετροϊκής θεραπείας υπό την παρουσία ελλειπουσών τιμών λόγω διαφορετικών μηχανισμών αποκοπής των ασθενών)

Christos Thomadakis
Biostatistician-Mathematician

Doctoral dissertation

Athens, 2022

HELLENIC REPUBLIC

National and Kapodistrian University of Athens

EST. 1837

National and Kapodistrian University of Athens

School of Health Sciences, School of Medicine

Sector of Social Medicine, Psychiatry, Neurology

Department of Hygiene, Epidemiology and Medical Statistics

Director: Pagona Lagiou

Statistical modeling of CD4 cell count evolution in HIV-positive individuals before and after antiretroviral treatment initiation under missing data due to various dropout mechanisms

(Μοντέλα διαχρονικής εξέλιξης CD4 λεμφοκυττάρων σε HIV οροθετικά άτομα πριν και μετά την έναρξη αντιρετροϊκής θεραπείας υπό την παρουσία ελλειπουσών τιμών λόγω διαφορετικών μηχανισμών αποκοπής των ασθενών)

Christos Thomadakis
Biostatistician-Mathematician

Doctoral dissertation

Athens, 2022

Στοιχεία διδακτορικής διατριβής

Ημερομηνία αιτήσεως: 08/01/2016

Ημερομηνία ορισμού 3μελούς Συμβουλευτικής Επιτροπής: 23/03/2016

Τα μέλη της 3μελούς Συμβουλευτικής Επιτροπής:

Γιώτα Τουλούμη, Καθηγήτρια (Επιβλέπουσα)

Λουκία Μελιγκοτσίδου, Αναπληρώτρια Καθηγήτρια (Μέλος)

Constantin T. Yiannoutsos, Professor (Μέλος)

Ημερομηνία ορισμού του Θέματος: 15/02/2017

Ημερομηνία καταθέσεως της διδακτορικής διατριβής: 01/04/2022

Πρόεδρος Ιατρικής Σχολής ΕΚΠΑ

Καθηγητής Γεράσιμος Δ. Σιάσος

Μέλη 7μελούς Εξεταστικής Επιτροπής:

Γιώτα Τουλούμη, Καθηγήτρια (Επιβλέπουσα)

Constantin T. Yiannoutsos, Professor

Λουκία Μελιγκοτσίδου, Αναπληρώτρια Καθηγήτρια

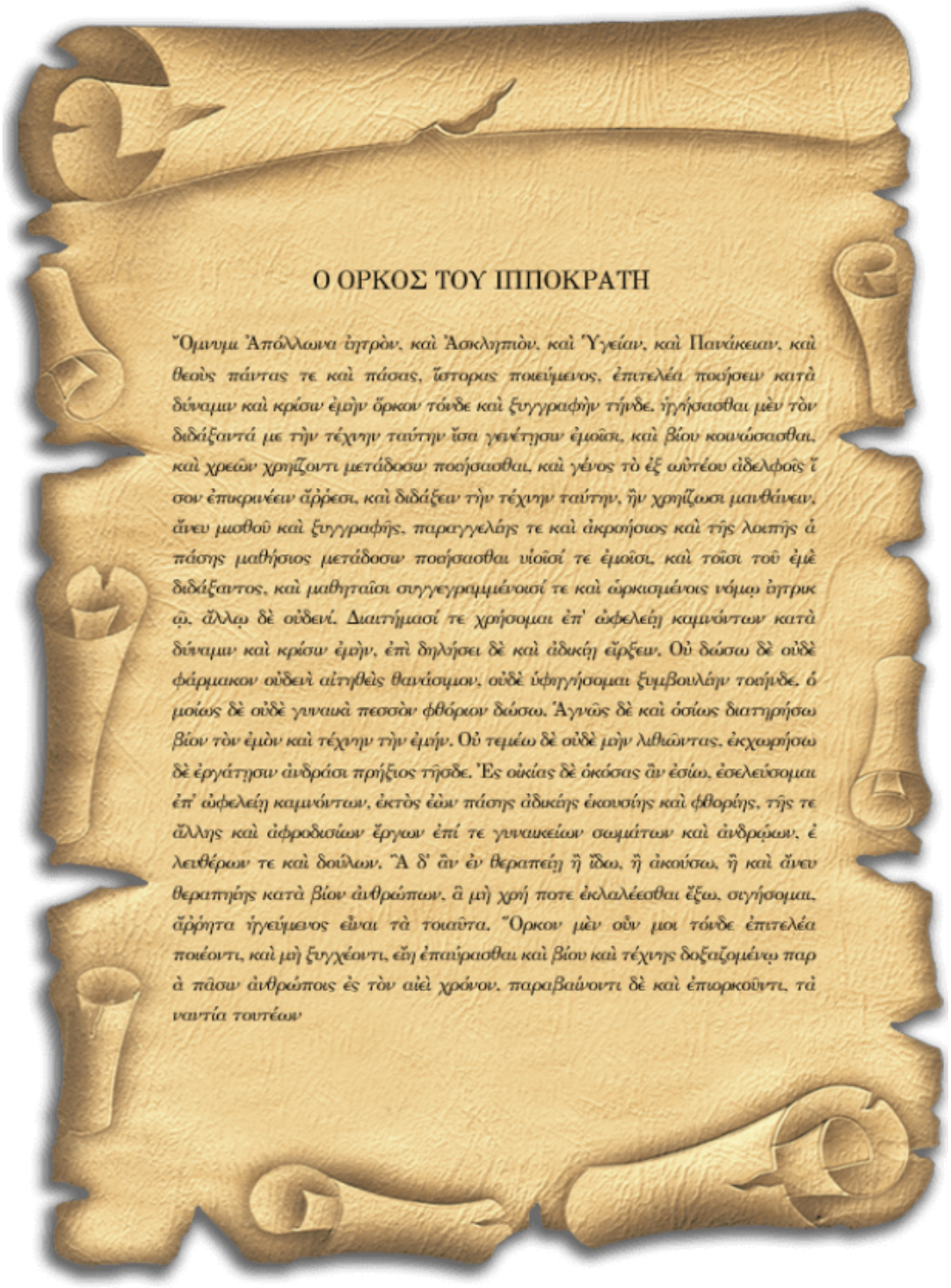
Βάνα Σύψα, Αναπληρώτρια Καθηγήτρια

Δεμίρης Νικόλαος, Επίκουρος Καθηγητής

Φώτιος Σιάννης, Επίκουρος Καθηγητής

Πανταζής Νικόλαος, Επίκουρος Καθηγητής

Η Διατριβή έγινε αποδεκτή με βαθμό ΑΡΙΣΤΑ



Ο ΟΡΚΟΣ ΤΟΥ ΙΠΠΟΚΡΑΤΗ

Ὄμνυμι Ἀπόλλωνα ἰητρὸν, καὶ Ἀσκληπιὸν, καὶ Ὑγείαν, καὶ Πανάκειαν, καὶ θεοὺς πάντας τε καὶ πάσας, ἵστορας ποιεύμενος, ἐπιτελέα ποιήσειν κατὰ δύναμιν καὶ κρίσιν ἐμὴν ὅρκον τόνδε καὶ ξυγγραφὴν τήνδε. ἡγήσασθαι μὲν τὸν διδάξαντά με τὴν τέχνην ταύτην ἴσα γενέτῃσιν ἐμοῖσι, καὶ βίου κοινώσασθαι, καὶ χρεῶν χρηίζοντι μετάδοσιν ποιήσασθαι, καὶ γένος τὸ ἐξ αὐτέου ἀδελφοῖς ἴσον ἐπικρινέειν ἄῤῥέσι, καὶ διδάξειν τὴν τέχνην ταύτην, ἢν χρηίζωσι μανθάνειν, ἄνευ μισθοῦ καὶ ξυγγραφῆς, παραγγελίης τε καὶ ἀκροήσιος καὶ τῆς λοιπῆς ἁπάσης μαθήσιος μετάδοσιν ποιήσασθαι υἱοῖσί τε ἐμοῖσι, καὶ τοῖσι τοῦ ἐμὲ διδάξαντος, καὶ μαθηταῖσι συγγεγραμμένοισί τε καὶ ὡρκισμένοις νόμῳ ἰητρικῷ, ἄλλῳ δὲ οὐδενί. Διαιτήμασί τε χρήσομαι ἐπ' ὠφελείῃ καμνόντων κατὰ δύναμιν καὶ κρίσιν ἐμὴν, ἐπὶ δηλήσει δὲ καὶ ἀδικίῃ εἴρξειν. Οὐ δώσω δὲ οὐδὲ φάρμακον οὐδενὶ αἰτηθεὶς θανάσιμον, οὐδὲ ὑφηγήσομαι ξυμβουλὴν τοιήνδε. ὁμοίως δὲ οὐδὲ γυναικὶ πεσσὸν φθόριον δώσω. Ἁγνῶς δὲ καὶ ὁσίως διατηρήσω βίον τὸν ἐμὸν καὶ τέχνην τὴν ἐμήν. Οὐ τεμέω δὲ οὐδὲ μὴν λιθιῶντας, ἐκχωρήσω δὲ ἐργάτῃσιν ἀνδράσι πρήξιος τῆσδε. Ἐς οἰκίας δὲ ὁκόσας ἂν ἐσίω, ἐσελεύσομαι ἐπ' ὠφελείῃ καμνόντων, ἐκτὸς ἐὼν πάσης ἀδικέης ἑκουσίης καὶ φθορίης, τῆς τε ἄλλης καὶ ἀφροδισίων ἔργων ἐπί τε γυναικείων σωμάτων καὶ ἀνδρώων, ἐ λευθέρων τε καὶ δούλων. Ἃ δ' ἂν ἐν θεραπείῃ ἢ ἴδω, ἢ ἀκούσω, ἢ καὶ ἄνευ θεραπηίης κατὰ βίον ἀνθρώπων, ἃ μὴ χρή ποτε ἐκλαλέεσθαι ἔξω, σιγήσομαι, ἄῤῥητα ἡγεύμενος εἶναι τὰ τοιαῦτα. Ὅρκον μὲν οὖν μοι τόνδε ἐπιτελέα ποιέοντι, καὶ μὴ ξυγχέοντι, εἴη ἐπαύρασθαι καὶ βίου καὶ τέχνης δοξαζομένῳ παρ' ἀπάσιν ἀνθρώποις ἐς τὸν αἰεὶ χρόνον. παραβαίνοντι δὲ καὶ ἐπιορκοῦντι, τἀναντία τουτέων

Η έγκριση της διδακτορικής διατριβής από την Ιατρική Σχολή του Πανεπιστημίου Αθηνών δεν αποτελεί παραδοχή των γνωμών του συγγραφέως. Ν5343/32, άρθρο 202, παράγραφος 2

Curriculum Vitae

Personal Information

Name	Christos Thomadakis
Date of Birth	24/02/1989.
Place of Birth	Athens, Greece
E-mail	cthomadak@med.uoa.gr

Professional Experience

2015–Current	**Biostatistician Research Associate** Department of Hygiene, Epidemiology and Medical Statistics, Department of Medicine, School of Health Sciences
2015–Current	**Lab teaching assistant** Teaching assistant for the "Survival Analysis", "Longitudinal data analysis", and "Clinical Trials" courses in the Postgraduate Program of Studies in Biostatistics at the National and Kapodistrian University of Athens

Education

2006–2012	Bachelor in Mathematics, National and Kapodistrian University of Athens (Final grade: 7.40)
2012–2015	MSc in Biostatistics National and Kapodistrian University of Athens (Final grade: 9.57).

Honours and awards

2016	**Student Conference Award Winner** Annual Conference of the International Society of Clinical Biostatistics (ISCB), Birmingham, 2016

Student conference award from the student conference awards committee of ISCB 2016 for the paper entitled "Longitudinal and time-to-drop-out joint models can lead to seriously biased estimates when the drop-out mechanism is at random"

2017 **Student Competition Winner**

Conference of the Eastern Mediterranean Region of the International Biometrics Society (EMR-IBS)

Student competition award from the scientific committee of EMR-IBS for the paper entitled "Misspecifying the covariance structure in a linear mixed model under MAR dropout"

2017 **Best Poster Award Winner**

Annual Conference of the International Society of Clinical Biostatistics (ISCB), Vigo, 2017

Best poster award from the scientific program committee of ISCB 2017 for the poster presentation "Comparing the fit of linear mixed models with different covariance structures based on Bayesian model comparison"

2020 **Research grant from Gilead Hellas**

Project "Immunologic and clinical response of people living with HIV on antiretroviral treatment: consequences of disengagement from care"

Publications

2016 Papatheodoridis, G., Vlachogiannakos, I., Cholongitas, E., Wursthorn, K., **Thomadakis, C.**, Touloumi, G. and Petersen, J. (2016), Discontinuation of oral antivirals in chronic hepatitis B: A systematic review. *Hepatology*

2017 Gournellis, R, Tournikioti, K, Touloumi, G, **Thomadakis, C**, Michalopoulou, PG, Christodoulou, C, Papadopoulou, A, Douzenis, A. (2017) Psychotic (delusional) depression and suicidal attempts: a systematic review and meta-analysis. *Acta Psychiatrica Scandinavica*

2017 Alvarez-del Arco D, Fakoya I, **Thomadakis C**, Pantazis N, Touloumi G, et al (2017), Application of a Bayesian approach to estimate the posterior probability of post-migration HIV acquisition for migrants participating in the Advancing Migrant Access to Health Services in Europe (aMASE) study. *AIDS*

2018 Skafida A, Mitrakou A, Georgiopoulos G, Alevizaki M, Spengos K, Takis K, Ntaios G, **Thomadakis C**, Vemmos K. (2018), In-hospital dynamics of glucose, blood pressure and temperature predict outcome in patients with acute ischaemic stroke. *European Stroke Journal*

2018 Gournellis, R, Tournikioti, K, Touloumi, G, **Thomadakis, C**, et al (2018), Psychotic (delusional) depression and completed suicide: a systematic review and meta-analysis. *Ann Gen Psychiatry*

2018 **Thomadakis, C**, Meligkotsidou, L, Pantazis, N, Touloumi, G. (2018), Longitudinal and time-to-drop-out joint models can lead to seriously biased estimates when the drop-out mechanism is at random. *Biometrics*

2019 Pantazis N, **Thomadakis C**, del Amo J, Alvarez-del Arco D, Burns FM, Fakoya I, Touloumi G (2019) Determining the likely place of HIV acquisition for migrants in Europe combining subject-specific information and biomarkers data. Statistical Methods in Medical Research

2020 **Thomadakis, C**, Meligkotsidou, L, Pantazis, N, Touloumi, G. (2020), Misspecifying the covariance structure in a linear mixed model under MAR drop-out. *Statistics in Medicine*

2021 **Thomadakis, C**, Meligkotsidou, L, Pantazis, N, Touloumi, G. (2021), Rejoinder to "Biased Estimation With Shared Parameter Models in the Presence of Competing Dropout Mechanisms". *Biometrics*

2021 Touloumi, G., **Thomadakis C** et al. (2022), HIV continuum of care: bridging cross-sectional and longitudinal analyses. *AIDS*

Submitted papers

2022 **Thomadakis, C**, Yiannoutsos, CT et al
The effect of HIV treatment interruption on subsequent immunologic response. Under revision in *American Journal of Epidemiology*

2022 **Thomadakis, C**, Meligkotsidou, L, Yiannoutsos, CT, Touloumi, G. Joint modeling of longitudinal and competing-risks data using cumulative incidence functions for the failure submodels accounting for potential failure cause misclassification through double sampling. Under revision in *Biostatistics*

2022 **Thomadakis, C**, Pantazis, N, Touloumi, G.
Issues with the expected information matrix of linear mixed models provided by popular statistical packages under MAR dropout. Submitted to *Statistics in Medicine*

Oral presentations in conferences

2016	**20th International Workshop on HIV Observational Databases (IWHOD)** Determining the likely place of HIV acquisition for migrants in Europe combining information on migration history, risk factors and biomarkers data.
2016	**37th Annual Conference of the International Society of Clinical Biostatistics (ISCB)** Longitudinal and time-to-drop-out joint models can lead to seriously biased estimates when the drop-out mechanism is at random
2017	**9th Conference of the Eastern Mediterranean Region of the International Biometrics Society (EMR-IBS)** Misspecifying the covariance structure in a linear mixed model under MAR dropout
2018	**XXIX International Biometric Conference (IBC)** Serious bias in competing risks shared parameter joint models when at least one of the failure types is at random
2019	**40th Annual Conference of the International Society of Clinical Biostatistics (ISCB)** Joint modeling of longitudinal and competing-risks data using cumulative incidence functions for the failure submodels
2020	**International Biometric Society (IBS) Virtual Journal Club** Longitudinal and time-to-drop-out joint models can lead to seriously biased estimates when the drop-out mechanism is at random (invited virtual presentation)
2021	**11th Conference of the Eastern Mediterranean Region of the International Biometrics Society (EMR-IBS)**

Joint modeling of longitudinal and competing-risks data using cumulative incidence functions for the failure submodels accounting for potential failure cause misclassification through double sampling (invited virtual presentation)

Abstract

In longitudinal epidemiological studies, the evolution of markers related to disease progression is often of great interest. However, longitudinal marker data often suffer from the presence of missing data, frequently in the form of dropout. Missing data though can ultimately undermine inferences under certain cases. It has been shown that likelihood-based methods, such as linear mixed models (LMMs) could provide unbiased results even after ignoring the dropout mechanism, given that the dropout probabilities depend on the observed marker values (MAR). When missingness depends on unobserved quantities (MNAR), joint modeling of the marker evolution and the dropout process is generally required for unbiased estimates. In this thesis, we address several issues arising due to incomplete data, motivated by the epidemiology of HIV, focusing mainly on longitudinal modeling of CD4 counts before/after treatment initiation.

When modeling CD4 count trajectories during untreated HIV infection, CD4 counts are mainly censored due to treatment initiation, with the nature of this mechanism remaining debatable (MAR or MNAR). Several shared-random effects models (SREMs), a specific subclass of MNAR joint models, have been fitted to such data. Motivated by this example, we analytically show that specific SREMs, when fitted to data subject to certain MAR mechanisms, can produce seriously biased marker rate of change estimates. In addition, we propose a more robust alternative SREM model that works well under specific MAR and MNAR dropout mechanisms.

Under MAR, CD4 cell counts during untreated HIV infection are usually

modeled through LMMs with random intercept and random slope. We analytically show that using a random intercept and slope structure when the true covariance structure is more complex can lead to seriously biased estimates, with the degree of bias depending on the magnitude of the MAR drop-out. Under misspecified covariance structure, we compare, in terms of induced bias, the approach of adding a fractional Brownian motion (BM) process on top of random intercepts and slopes with the approach of using splines for the random effects. Moreover, to discriminate between the examined approaches in real-data applications, we adopt Bayesian model comparison criterion based on the posterior model probabilities.

When modeling CD4 counts after treatment initiation, dropout can occur due to death in care or disengagement from care, which are competing events. Death is usually considered to correspond to MNAR dropout though the nature of the mechanism of disengagement from care is less clear. This setting calls for joint modeling. We propose a flexible class of SREMs to jointly model the marker evolution and multiple causes of failure using cumulative incidence functions (CIFs) in the survival submodels, with the CIFs depending on the "true" marker value over time. The fact that the all-cause CIF should be bounded by 1 is formally considered. The proposed models are extended to account for potential failure cause misclassification through double sampling. We also provide a multistate representation of the whole population by defining mutually exclusive discrete states based on the "true" marker values and the competing risks. Based solely on the assumed joint model, we derive fully Bayesian inference for state occupation and transition probabilities.

Σε διαχρονικές επιδημιολογικές μελέτες, η εξέλιξη βιο-δεικτών που σχετίζονται με την εξέλιξη της νόσου είναι συχνά κύριου ενδιαφέροντος. Ωστόσο, σε διαχρονικά δεδομένα δεικτών συχνά δημιουργούνται δυσκολίες από την παρουσία ελλειπουσών τιμών, συνήθως στη μορφή της περικοπής. Ελλείποντα δεδομένα μπορούν όμως να υπονομεύσουν τα συμπεράσματα σε συγκεκριμένες περιπτώσεις. Έχει αποδειχθεί ότι μέθοδοι βασιζόμενες στην πιθανοφάνεια, όπως τα μεικτά γραμμικά μοντέλα, μπορούν να αποδώσουν αμερόληπτες εκτιμήσεις αγνοώντας το μηχανισμό περικοπής δεδομένου ότι οι πιθανότητες περικοπής εξαρτώνται από τις παρατηρηθείσες τιμές του δείκτη (τυχαίος μηχανισμός περικοπής). Όταν οι πιθανότητες περικοπής εξαρτώνται από μη παρατηρούμενες ποσότητες (μη τυχαίος μηχανισμός περικοπής), από κοινού μοντελοποίηση της εξέλιξης του δείκτη και του μηχανισμού περικοπής χρειάζεται συχνά για αμερόληπτες εκτιμήσεις. Στην παρούσα διατριβή, αντιμετωπίζουμε πολλά ζητήματα που προκύπτουν λόγω ημιτελών δεδομένων, ορμώμενοι από την επιδημιολογία του ιού HIV, εστιάζοντας κυρίως στη διαχρονική μοντελοποίηση των CD4 λεμφοκυττάρων πριν και μετά την έναρξη θεραπείας.

Όταν μοντελοποιούνται οι πορείες του αριθμού των CD4 λεμφοκυττάρων, οι μετρήσεις των CD4 λεμφοκυττάρων περικόπτονται λόγω έναρξης θεραπείας, με τη φύση του συγκεκριμένου μηχανισμού να παραμένει αμφίβολη. Αρκετά μοντέλα κοινών τυχαίων επιδράσεων [shared-random effects models (SREMs)], μια συγκεκριμένη υποκατηγορία μη τυχαίων από κοινού μοντέλων, έχουν εφαρμοστεί σε τέτοια δεδομένα. Παρακινούμενοι από αυτό το παράδειγμα, δείχνουμε αναλυτικά ότι συγκεκριμένα SREM μοντέλα, όταν εφαρμοστούν σε ημιτελή σύνολα δεδομένα λόγω ενός τυχαίου μηχανισμού, μπορεί να αποδώσουν σημαντικά μεροληπτικές εκτιμήσεις για το ρυθμό μεταβολής του δείκτη. Επιπρόσθετα, προτείνουμε ένα πιο ανθεκτικό εναλλακτικό SREM μοντέλο το οποίο αποδίδει καλά κάτω από συγκεκριμένους τυχαίους και μη τυχαίους μηχανισμούς περικοπής.

Υποθέτοντας τυχαίο μηχανισμό περικοπής, ο αριθμός των CD4 λεμφοκυττάρων

κατά τη διάρκεια της HIV λοίμωξης χωρίς θεραπεία μοντελοποιείται μέσω μεικτών γραμμικών μοντέλων τυχαίας σταθεράς και κλίσης. Αποδεικνύουμε αναλυτικά ότι μια δομή τυχαίας σταθεράς και κλίσης, όταν η πραγματική δομή συνδιακύμανσης είναι πιο περίπλοκη, μπορεί να οδηγήσει σε σοβαρά μεροληπτικές εκτιμήσεις, με το βαθμό της μεροληψίας να εξαρτάται από την ένταση του τυχαίου μηχανισμού περικοπής. Υπό λανθασμένη δομή συνδιακύμανσης, συγκρίνουμε, με βάση την επαγόμενη μεροληψία, την προσέγγιση της πρόσθεσης μιας κλασματικής κίνησης Brown διαδικασίας [fractional Brownian motion (BM)] σε ένα μοντέλο τυχαίας σταθεράς και κλίσης με την προσέγγιση της χρησιμοποίησης κατά τμήματα πολυωνυμικών συναρτήσεων (splines) για τις τυχαίες επιδράσεις. Επίσης, για να επιλέξουμε μεταξύ των υπό εξέταση προσεγγίσεων σε εφαρμογές με πραγματικά δεδομένα, υιοθετούμε ένα κριτήριο Μπεϋζιανής σύγκρισης μοντέλου βάσει των εκ των υστέρων πιθανοτήτων των μοντέλων.

Όταν μοντελοποιούνται οι μετρήσεις CD4 μετά την έναρξη θεραπείας, περικοπή μπορεί να συμβεί λόγω θανάτου κατά τη διάρκεια της φροντίδας ή απόσυρσης από τη φροντίδα, τα οποία ενδεχόμενα είναι ανταγωνιστικοί κίνδυνοι. Ο θάνατος συνήθως θεωρείται ότι αντιστοιχεί σε μη τυχαία περικοπή, ενώ η φύση του μηχανισμού απόσυρσης από τη φροντίδα είναι λιγότερο ξεκάθαρη. Σε ένα τέτοιο παράδειγμα ενδείκνυται από κοινού μοντελοποίηση. Προτείνουμε μια ευέλικτη υποκατηγορία SREM μοντέλων για την από κοινού μοντελοποίηση πολλαπλών αιτιών αποτυχίας μέσω των συναρτήσεων αθροιστικής επίπτωσης στα υπο-μοντέλα των αιτιών αποτυχίας, με τις συναρτήσεις αθροιστικής επίπτωσης να εξαρτώνται από την «πραγματική» τιμή του δείκτη στο χρόνο. Το γεγονός ότι η συνολική αθροιστική επίπτωση πρέπει να είναι μικρότερη του 1 λαμβάνεται ρητά υπόψη. Η προτεινόμενη μεθοδολογία έχει επεκταθεί για να λάβει υπόψη πιθανή δυσταξινόμηση των αιτιών αποτυχίας μέσω διπλής δειγματοληψίας. Παρέχουμε επίσης μια αναπαράσταση, μέσω πολλαπλών καταστάσεων, του συνολικού πληθυσμού ορίζοντας αμοιβαία αποκλειόμενες καταστάσεις βάσει των

«πραγματικών» τιμών του δείκτη και των ανταγωνιστικών κινδύνων. Βασιζόμενοι αποκλειστικά στο από κοινού μοντέλο, εξάγουμε πλήρως Μπεϋζιανή συμπερασματολογία για πιθανότητες καταστάσεων και πιθανότητες μεταβάσεων μεταξύ καταστάσεων.

Στην οικογένεια μου και τους φίλους μου

Acknowledgements

Θα ήθελα να ευχαριστήσω την τριμελή επιτροπή της διατριβής και ιδιαιτέρως την επιβλέπουσα Καθηγήτρια Γιώτα Τουλούμη καθώς και την Αναπληρώτρια Καθηγήτρια Λουκία Μελιγκοτσίδου για την υποστήριξη, τις συμβουλές και τις πολύτιμες εμπειρίες που μου προσφέρθηκαν στο πλαίσιο της διατριβής. Επίσης, θα ήθελα να ευχαριστήσω τον Επίκουρο Καθηγητή Νίκο Πανταζή για τη σημαντική βοήθεια του κατά τα χρόνια διεξαγωγής της διατριβής.

Contents

List of Figures

List of Tables

LIST OF TABLES

Glossary

$()^\top$	Transpose of a matrix vector
$()_0$	Denotes the true parameter values of the corresponding parameter
a_{0j}	True marginal probability of $M = j$
$\boldsymbol{\alpha}$	Association parameter of a proportional hazards SREM model
$\boldsymbol{b}$	Random effects vector
$\boldsymbol{\beta}$	Fixed effects vector
$\boldsymbol{D}$	Covariance matrix of the random effects
$\boldsymbol{\epsilon}$	Vector of within-subjects residuals at the scheduled measurement times
$\boldsymbol{\lambda}$	Association parameters (relating to the random effects) of a lognormal model
$\boldsymbol{\theta}$	Parameter vector of a joint model of longitudinal and survival data
$\boldsymbol{\theta}_L$	Parameter vector of the marker model
$\boldsymbol{\theta}_t$	Parameter vector of the dropout/missing data mechanism
$\boldsymbol{\theta}_v$	Parameter vector of variance components
$\boldsymbol{V}_{(j)}$	Marginal covariance matrix of $\boldsymbol{Y}_{(j)}$
$\boldsymbol{X}$	Design matrix for the fixed effects at the scheduled measurement times
$\boldsymbol{X}_i(t)$	Fixed effects design matrix for the ith individual at time t
$\boldsymbol{Y}$	Marker measurements intended to be collected
$\boldsymbol{Y}_{(\bar{j})}$	Missing marker measurements
$\boldsymbol{Y}_{(j)}$	Observed marker marker measurements
$\boldsymbol{Z}$	Design matrix for the random effects at the scheduled measurement times

GLOSSARY

$\boldsymbol{Z}_i(t)$ Random effects design matrix for the ith individual at time t

K_i True failure cause in competing risks models

$\tilde{K}_i$ Observed, and potentially misclassified, failure cause in competing risks models

$\mathcal{I}$ Generic notation for the information matrix of any model

$\mathcal{U}$ Generic notation for the score vector of any model

ω Within-subjects precision ($1/\sigma^2$)

σ^2 Within-subjects variance

$\boldsymbol{\theta}_{misc}$ Parameter vector associated with the probability of misclassifying the true failure cause

ζ Mean of a lognormal model

$h(t)$ Hazard function (for continuous time)

$h_0(t)$ Baseline hazard function

M Number of the observed marker measurements

$m_i(t) = \boldsymbol{X}_i^\top(t)\boldsymbol{\beta} + \boldsymbol{Z}_i^\top(t)\boldsymbol{b}_i$ "True" marker value for the ith individual at time t

r Residuals of a lognormal model

s^2 Variance of a lognormal model

$t_1, t_2, \ldots, t_Q$ Scheduled measurement times

U Log time to dropout

GEE Generalized estimating equation

LMM Linear mixed model

LN-SREM(RE) Lognormal SREM joint model

MAR Missing at random

MCAR Missing completely at random

MNAR Missing not at random

PH-SREM(CV) Proportional hazards SREM joint model depending on the current "true" marker value

PH-SREM(LV,RE) Proportional hazards SREM joint model depending on both the last observed marker value and the random effects

PH-SREM(RE) Proportional hazards SREM joint model depending on the random effects

Q Maximum number of measurements

SREM Shared random effects model

SREM-CIF-1 Competing risks SREM using proportional subdistribution hazards submodels

SREM-CIF-2 Competing risks SREM using the generalized odds rate transformation of CIFs for the failure submodels

VL HIV viral load

Chapter 1

General Introduction

1.1 Statistical modeling in longitudinal studies

In many studies in various disciplines, data can be repeatedly collected over time on the same individuals (referred to as longitudinal data), or the data can be collected at different points in space, leading to spatial data. In this work, we focus on the former case, with the corresponding studies being referred to as longitudinal studies. Typically, in longitudinal studies, information on many variables/characteristics is recorded at different visits on the same individuals, which makes the study of within-individual evolution feasible. In longitudinal epidemiological studies concerning chronic diseases, the evolution of biochemical markers that are related to disease progression is often of primary interest. The advantages of using such markers in tracking disease progression are comprehensibly described in Jewell and Kalbfleisch (1992). In contrast to longitudinal studies, in cross-sectional ones, data on a single time point are collected for each individual though the time points differ between individuals. Thus, inferences can be obtained only about the marginal marker evolution of the population, but studying the within-individual correlation is impossible. On the other hand, longitudinal studies can explicitly model the correlation and are expected to be more powerful than the corresponding cross-sectional ones as, for most outcomes, the variability across individuals, e.g. due to unmeasured characteristics such as genetic or environmental factors, is sub-

stantially greater than the within-individual variability. The effects of such factors are canceled out in longitudinal studies as within-individual marker changes are ultimately estimated, with each individual considered to serve as his/her own control. Also, longitudinal studies can distinguish the age and cohort effects, i.e. the direct effects of aging and the effects implied by comparing groups of individuals with different baseline ages, respectively (Diggle et al., 2002).

Since marker data from the same individuals tend to be (positively) correlated, most statistical approaches for modeling longitudinal data take that correlation into account. It should be emphasized, though, that the correlation is not usually of scientific interest, except for special cases (e.g. in genetic or environmental studies). Some methods explicitly postulate a model for the covariance structure of the data, such as compound symmetry, exponential correlation, Toeplitz, or unstructured covariance (e.g. Diggle et al., 2002, Fitzmaurice et al., 2011). Parameter estimation can be carried out by employing either a fully parametric model (e.g. using the Normal distribution) or a semi-parametric model using the generalized estimating equation (GEE) approach, initially introduced by Liang and Zeger (1986).

Another popular class of models for describing longitudinal data is known as random-effect models. In these models, the covariance structure is specified by using subject-specific latent quantities referred to as the random effects, firstly introduced by Harville (1977) and popularized by Laird and Ware (1982), though these ideas derive from the ANOVA paradigm. The random effects are considered to be inherent characteristics of individuals, assumed to remain constant over time, and assigned a prior distribution, e.g. the Normal distribution. Hence, after integrating out the random effects, a specific model for the marginal covariance structure of the data is implicitly assumed. One frequently applied random-effect model is the random intercept and slope model, which is more suitable when the marker evolves more or less linearly over time. This model assumes that the baseline marker value is subject-specific and it is equal to a population-averaged baseline plus a "random intercept" term, measuring the deviation of the individual's baseline marker value from the population baseline. The rate of

change is assumed to be linear and subject-specific, with the "random slope" measuring the difference of the individual's rate of change relative to the population-averaged rate of change.

In most longitudinal studies, in principle, the protocol assumes a regular time schedule for collecting data, i.e. measurements are expected to be recorded at prespecified fixed times. Although fixed visit times can greatly facilitate data collection and statistical analysis, this is rarely the case in studies involving human subjects, resulting in highly imbalanced data and different numbers of observations among individuals. Some reasons for the occurrence of imbalanced data include staggered entry of individuals, missed visits, death, and dropout of the study based on individual's decision. Most methods can easily handle imbalanced longitudinal data (Harville, 1977, Harville and Mee, 1984, Laird and Ware, 1982, Liang and Zeger, 1986, Lindstrom and Bates, 1988, Zeger and Karim, 1991, Zeger et al., 1988). For a more comprehensive review of the available methods for longitudinal data analysis, we refer to Fitzmaurice et al. (2008).

Apart from imbalanced data, there are many types and reasons for missing data. In general, the pattern of missingness may be either monotone (the individual drops permanently out of the study, often referred to as attrition or dropout) or intermittent (missing observations between the observed ones). In epidemiological longitudinal studies, it is commonplace to collect data on time to event outcomes apart from the longitudinal marker data. In many cases, the occurrence of the event of interest (e.g. death) precludes subsequent data collection on the marker, leading to monotone missingness. It should be also noted that in some instances, the occurrence of an event may influence the marker distribution so drastically that the marker measurements collected after the event have to be excluded from the analysis. This is also considered to be a missing data problem even though there are no actually missing data, strictly speaking. A classic example occurs in the pre-marketing stages of drug development, where the focus is often on drug efficacy rather than effectiveness, with observations collected after study drug discontinuation excluded from the analysis, following the so-called "de jure" estimand. Irrespective of whether the data are actually missing or excluded from

the statistical analysis, missing data can have serious consequences on the analyses.

Research on the quality of inferences drawn from incomplete data has grown since the seminal paper by Rubin (1976). The main implication from Rubin (1976) is that the validity of inferences depends crucially on the nature of the missing data mechanism, i.e. the probability of a marker value being missing conditional on the marker values that are intended to be collected. The taxonomy made by Rubin (1976), which is widely adopted in the literature, categorizes the missing data mechanisms into three groups: (i) missing completely at random (MCAR), (ii) missing at random (MAR), and (iii) missing not at random (MNAR). Data are said to be MCAR when the missingness probabilities are unrelated to all marker measurements that are intended to be collected. This mechanism poses no problems to the analyses apart from a higher variability due to lower sample size. A standard example of MCAR data is when individuals drop out of the study due to work or other considerations that are unrelated to the individual's health status. Data are said to be MAR when the missingness probabilities depend on the set of the observed marker measurements, but are unrelated to the marker values that would have been observed. A standard example of MAR data occurs when some information is collected only when the current marker value exceeds or drops below some predetermined threshold. Finally, a missing data mechanism is MNAR when the probability of a marker measurement being missing depends on the marker values that should have collected, in addition to the ones actually observed. This is the most difficult case to deal with.

Based on standard theory (Laird, 1988, Little and Rubin, 2002, Rubin, 1976), likelihood-based methods, based on either the Frequentist or Bayesian approach modeling all observed marker data, provide unbiased estimates given that the correct model is applied and the missing data mechanism is MAR. This, in turn, implies that if the fitted model does not adequately describe the true covariance structure of the data, the estimates for the mean marker evolution over time can be biased under MAR missingness. An additional implication is that approaches that are not based on the likelihood function, e.g. semi-parametric methods such as the GEE models, can produce biased

estimates under MAR data. In any case, though, such biases can be easily resolved by weighting the GEE estimator by the inverse of the estimated probability of being observed (Robins et al., 1995), hence the term inverse probability weighting (IPW), which is frequently used in the literature. IPW as a general method of estimation was derived by Horvitz and Thompson (1952) and further developed by Robins et al. (1994), though extensions of IPW producing unbiased estimates under less restrictive assumptions have been proposed (Scharfstein et al., 1999). An alternative modeling framework to deal with missing longitudinal marker data is the multiple imputations approach proposed by Rubin (1987), in which the missing values are filled in, i.e. imputed, multiple times based on some stochastic mechanism. Then the model of interest is fitted to multiple "fully" observed datasets, combining the results using the so-called Rubin's rules (Rubin, 1987).

In general, the most difficult case to handle in practice is when the missing data mechanism is informative in the sense that the probability of missingness depends on values that are not (directly) observed (i.e. MNAR). Under MNAR, it has been shown that likelihood-based methods modeling the observed data only (i.e. ignoring the missing data mechanism) can lead to biased parameter estimates (Laird, 1988, Little and Rubin, 2002, Rubin, 1976). To fully eliminate such biases, one should jointly model the marker evolution over time and the missing data mechanism. Plenty of joint modeling approaches have been proposed, with the corresponding literature having become vast, including models with different aims and plenty of parameterizations. Some indicative references include (Diggle and Kenward, 1994, Faucett and Thomas, 1996, Gruttola and Tu, 1994, Huang et al., 2010, Little, 1993, 1994, Pantazis et al., 2005, Schluchter, 1992, Touloumi et al., 1999, Wu and Carroll, 1988, Wulfsohn and Tsiatis, 1997). The primary distinctive feature of these models constitutes the way the joint likelihood of the marker and missing data processes are factorized. In addition, much of the literature focuses on monotone missing data mechanisms in the form of dropout. Another difference between these models is about whether the primary motivation comes from the accurate estimation of the marker evolution over time adjusted for

informative missingness or from the accurate estimation of the risk for the dropout event conditionally on the "true" marker values. A special subclass of MNAR models which is frequently applied in biomedical settings is termed shared random effects models (SREMs) (e.g. Gruttola and Tu, 1994, Wulfsohn and Tsiatis, 1997) in which the marker evolution and the time to dropout are assumed to depend on the individual's random effects. In some SREMs (Gruttola and Tu, 1994, Schluchter, 1992, Touloumi et al., 1999), the time to dropout is directly modeled conditional on the random effects, where some parameters relate the random effects to the distribution of the time to dropout. Thus, these models can examine whether individuals with marker profiles worse than the population average are more likely to drop out of the study earlier. When the association parameters are estimated to be different from zero, the authors claimed that this is an indication of MNAR dropout. In other SREMs (e.g. Faucett and Thomas, 1996, Wulfsohn and Tsiatis, 1997), where the motivation is based on estimating the risk for dropout given the true marker value, a proportional hazards model is usually assumed for the time to dropout conditionally on the current "true" marker value.

In practice, there may be multiple reasons for dropout, usually in the form of competing risks, i.e. multiple mutually exclusive events. The most typical example of competing risks is death from different causes. One approach would be to combine all reasons into a composite one, but this is not generally recommended, especially when the marker is likely to affect the risk for each dropout event differentially. In the last few years, though, joint modeling of longitudinal data and competing risks time to event data has also gained attention (e.g. Andrinopoulou et al., 2014, 2017, Elashoff et al., 2008, Hickey et al., 2018, Hu et al., 2012, 2009, Huang et al., 2010, Proust-Lima et al., 2016, Williamson et al., 2008). In competing risks analysis, a frequent issue is that the failure cause is missing or potentially misclassified. To deal with this issue, a random sample from individuals with potentially misclassified or missing failure cause can be selected, with the true failure cause actively ascertained.

In applied epidemiological/medical research, especially for prediction purposes, pro-

gression of cohorts over time may be monitored by using mutually exclusive states defined jointly by competing risks data and discretized continuous marker data. The literature on multistate modeling is extensive but most approaches usually consider observed states (e.g Klein and Shu, 2002, Putter et al., 2007), with few methodological approaches on estimating multistate probabilities defined jointly by unobserved marker data and observed competing risks data (Hu et al., 2012).

1.2 Disease markers in people with HIV

The human immunodeficiency virus (HIV) is a retrovirus causing the acquired immune deficiency syndrome (AIDS), a condition in which the human immune system collapses, leading to the, so-called, opportunistic infections that in many cases result to death. HIV can be classified into two main types: HIV-1 and HIV-2. HIV-1 was discovered first and is more prevalent worldwide, while HIV-2 is less pathogenic and transmissible, and it is mostly confined to West Africa. So when we generally say HIV, we refer to HIV-1. The fundamental differences between HIV-1 and HIV-2 infections lie in the mechanism of retroviral pathogenesis, which has not been fully clear yet. There is large variability in the time from HIV infection to AIDS onset (Muñoz et al., 1989, Pantaleo et al., 1993, Phair et al., 1992), with the corresponding median time in the absence of treatment estimated to be around 10 years. However, due to the introduction of combined antiretroviral therapy (ART) in 1996, the mortality and morbidity of HIV has substantially decreased over time, tending to still improve in recent years. Therefore, if patients are diagnosed relatively early, initiate and remain compliant to treatment, they are at risk mainly for non-AIDS serious adverse events nowadays. Thus, HIV infection is now considered more as a chronic disease.

During this variable and often prolonged asymptomatic period after the acute viremic phase (Muñoz et al., 1989), several immunologic and virologic markers, such as the percentage of CD4 cell count, serum β_2-microglobulin and neopterin levels and HIV p24 antigenemia, have been considered as potential markers of disease progres-

sion (Fahey et al., 1990, Lang et al., 1989, Stein et al., 1992). Among all potential predictors, the most reliable ones are the absolute value of the CD4 cell count and the HIV-RNA viral load (VL). These markers have been routinely used to keep track of disease progression and have formed the basis of WHO guidelines for ART initiation for almost 20 years.

CD4 lymphocyte cells, also termed T-helper cells, are a subset of the total T lymphocytes. They are called helper cells because one of their main roles is to send signals to other types of immune cells, including CD8 killer cells, which then destroy the infectious particle. Thus, CD4 cells have a key role in the immune function but constitute the primary target cells of HIV. Many epidemiological studies have shown that the absolute number of CD4 cells gradually declines during the course of the disease (e.g. Detels et al., 1988, Eyster et al., 1987, Lang et al., 1989, Touloumi et al., 2012). More accurately, though, CD4 counts decrease very rapidly immediately after HIV infection (i.e. the acute HIV infection period). At approximately 6 weeks since infection, CD4 counts start to increase again up to around 12 weeks since infection, when HIV seroconversion occurs (i.e. the time point after which a patient can test positive for HIV). Then, in the absence of ART, the number of CD4 cells starts decreasing again but at a much slower rate. Many researchers have used transformations of CD4 counts (e.g. square root, log, fourth root) mainly to linearize their decline over time and normalize their distribution during the HIV natural history (disease evolution if untreated). It is also well known that a low count of CD4 cells is a reliable predictor for increased risk for development of AIDS, and that patients with steeper CD4 slopes tend to develop AIDS or die sooner. The performance of CD4 cell count as a surrogate measure for clinical events, such as AIDS onset or death, in clinical trials has also been evaluated in the literature (e.g. Choi et al., 1993, Tsiatis et al., 1992). Apart from evaluating the progression of the disease, CD4 decline during the HIV natural history is of paramount importance for other epidemiological purposes, as well, e.g. when interest focuses on predicting the time gap between HIV seroconversion and HIV diagnosis based on post-diagnosis CD4 counts (Pantazis et al., 2019b) or when estimating the HIV incidence

(van Sighem et al., 2015). CD4 cell count was the main factor for ART initiation based on the WHO guidelines from 1996 to 2015. That is, patients did not start ART immediately after diagnosis, but rather when the current CD4 count dropped below a certain cutoff, though this cutoff was regularly updated when new information became available. However, the results of the START study (Lundgren et al., 2015) in 2015 showed clearly the benefits of early initiation of ART in reducing drastically mortality and the risk for opportunistic infections. In addition, early initiation of ART leads to controlled viremia (i.e. undertectable HIV-RNA) which, in turn, minimizes the probability of transmission, having thus a great benefit not only for the individual but also for public health. Therefore, nowadays, ART is initiated soon after diagnosis, which implies that very few CD4 counts collected before ART initiation will be available from now on.

CD4 count recovery after ART initiation has also been thoroughly examined by many authors. Provided that patients adhere to treatment, a very rapid increase in CD4 counts is expected, though the probability of full CD4 recovery (e.g. CD4>800 cells/μL) depends to a large extent on the CD4 at ART initiation (e.g. Pantazis et al., 2019a, Stirrup et al., 2018) and on some other factors (Balestre et al., 2012), as well. Based on data from sub-Saharan Africa cohorts, though, it has been shown that the probability of sub-optimal immune recovery is considerable (Nakanjako et al., 2016).

Epidemiological research on VL, i.e. the amount of virus in blood, started to grow when reliable quantitative methods became available (Holodniy et al., 1991). Immediately after infection, HIV multiplies rapidly and a person's viral load is typically very high. After a few weeks to months, this rapid replication of HIV declines and the person's viral load drops to its set point and increase slowly afterwards. The interval between HIV seroconversion and viral set point is estimated to be around 1 year (Lyles et al., 2000, Touloumi and Hatzakis, 2000, Touloumi et al., 2004). What is essential is that the VL levels at viral set point are strongly associated with subsequent disease progression, and more importantly, independently of CD4 counts (Jurriaans et al., 1994, Mellors et al., 1995, 1996, O'Brien et al., 1996). The joint evolution of CD4 and VL

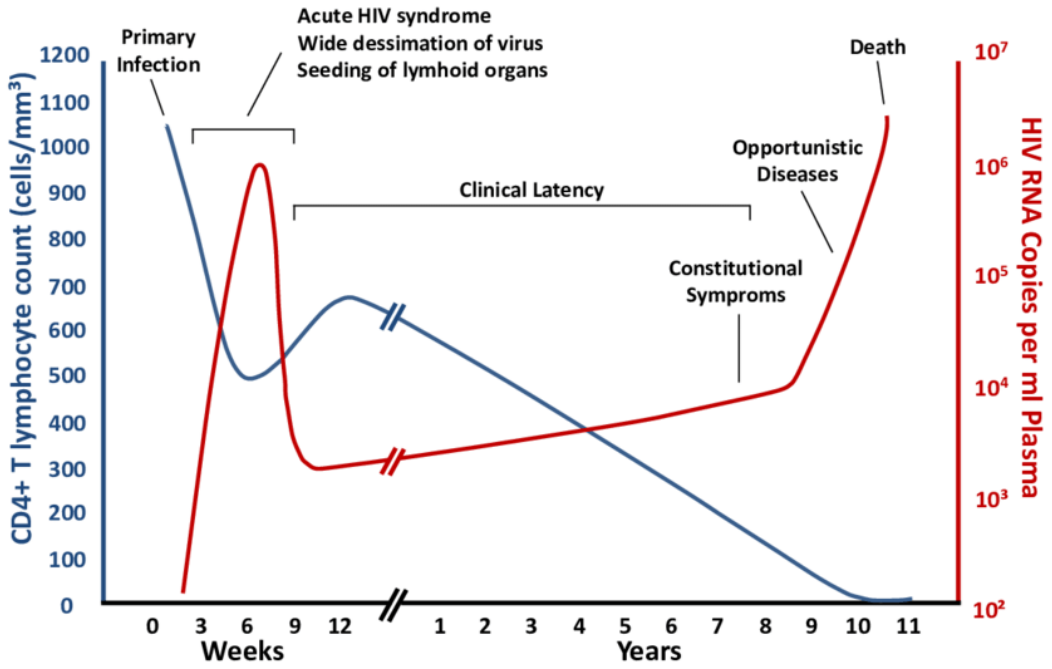


Figure 1.1: Generalised course of HIV infection. The evolution of the CD4 T cell count (blue) and viral load (red) during acute infection, clinical latency and AIDS phase (reproduced from Fauci AS et al. 1996 58).

evolution during the HIV natural history is described in Figure 1.1. After initiating ART, though, VL drops very steeply, by 90% within the first few days and by 99% within the first few weeks. Typically, in most people viral load become undetectable (i.e. VL result below detection limits) within one to three months given adherence to treatment.

1.3 HIV studies: The CASCADE and the East African IeDEA collaboration

We briefly describe the two databases used in this thesis:

The Concerted Action of Seroconversion to AIDS and Death in Europe (CASCADE) study (e.g. Collaboration, 2000, Pantazis et al., 2014, Touloumi et al., 2004) is a collaboration of HIV cohorts pooled within the European Coordinating Committee for the

Integration of Ongoing Coordination Actions Related to Clinical and Epidemiological HIV Research (EuroCoord; `www.EuroCoord.net`). All collaborating cohorts have received approval from their respective regulatory or national ethics review boards. The individual cohorts include only patients for whom the seroconversion date can be estimated with reasonable precision. For the majority of them, seroconversion dates were estimated as the midpoint between the last documented negative and first positive HIV antibody test dates with an interval of 3 years or less between tests. For the remainder, the seroconversion date was estimated through laboratory evidence of acute seroconversion or as the date of seroconversion illness. The purpose of CASCADE was to study HIV infection throughout the whole course of the disease, addressing research questions regrading both the HIV natural history and response to ART. The idea behind this collaboration was to address research questions that could not be reliably explored by the individual studies. Some descriptive characteristics of the CASCADE study data are provided in Sections 3.5 and 4.7.

The East African International Databases to Evaluate AIDS (IeDEA-EA) (Nakanjako et al., 2016, Tymejczyk et al., 2019) is one of the seven regional data centers funded by the U.S. National Institutes of Health to provide a rich resource for globally diverse HIV/AIDS data. Indiana University leads the East African region in collaboration with the University of California San Francisco (U.S.), Columbia University (U.S.), University of Toronto (Canada), Moi University (Kenya), Mbarara University (Uganda), the Kenyan Medical Research Institute (Kenya) and the Tanzanian National AIDS Control Program (Tanzania). The primary goal of this consortium is the provision of answers to questions that clinicians, programs, governments and international organizations consider central to the evolution and sustainability of their long term HIV care and treatment strategies. The current global focus is on achieving the UNAIDS 90-90-90 targets (Joint United Nations Programme on HIV/AIDS and Joint United Nations Programme on HIV/Aids, 2014). Descriptive characteristics of the included study participants are presented in Section 5.7.

1.4 Aims

Using the CASCADE and the IeDEA-EA studies as motivated examples, and the need to analyze the evolution of CD4 counts over time, we focus on two main aims: 1) to analyze CD4 counts changes over time prior to ART initiation (i.e., during natural history of the HIV infection) and 2) to analyze CD4 count trajectories after ART initiation. In both cases, longitudinal CD4 count measurements could be censored due to patient drop-out. The main mechanism of drop-out is ART initiation in the first case and death or disengagement from care in the second one.

For the first aim, we focus mainly on the nature of the drop-out mechanism. When the nature of the dropout mechanism is unknown, many researchers may apply SREMs, at least as part of a set of sensitivity analyses if informative dropout is suspected. Following standard recommendations from published papers (e.g. Gras et al., 2013, Schluchter, 1992, Touloumi et al., 1999), researchers may, perhaps wrongly, interpret deviations between results from LMMs (assuming ignorable dropout) and from SREMs as indicating MNAR missingness. The discrimination though between MAR and MNAR is difficult in practice (Molenberghs et al., 2008); thus one may end up fitting a SREM to data that are actually MAR. The aims of this part of the thesis are:

(a) to investigate the performance of SREMs when fitted to data subject to certain MAR mechanisms

(b) to propose a more robust model that could provide valid estimates in both MAR and MNAR cases, and

(c) to investigate the performance of the commonly used LMMs under misspecified covariance structure and to compare various approaches to identify the model with the "correct" covariance structure.

The various modeling approaches will be compared in simulation studies and fitted to CASCADE study data.

For the second aim, we focus on analyzing the evolution of a marker when more

than one dropout mechanisms are present. On top of that, we also examine the issue of potentially misclassified cause of dropout. Multiple dropout mechanisms (e.g. death and disengagement from care) yield competing-risk data, thus the issue here is to joint model longitudinal marker data along with competing-risk data under the SREM framework. The main aims of this part of the thesis are:

(a) to propose a unified and flexible approach to jointly model a continuous disease marker over time using CIFs for the failure (i.e., drop-out) submodels, and

(b) to propose a way to adjust for misspecified failure cause.

The methods developed/investigated in this part of the PhD will be examined in simulation studies and will be fitted to data from the IeDEA-EA cohort study, where clinical relevant predictions will be made.

1.5 Structure of the thesis

Chapter 2 presents an overview of the available methods for modelling longitudinal data with missingness. The notation used throughout this thesis and standard results from missing data theory are described first. Some frequently applied traditional approaches to modeling the observed longitudinal data solely, as well as commonly applied approaches to jointly modeling longitudinal data and the missing data process, are described in detail.

In Chapter 3, we examine the case in which SREM models are fitted to incomplete data due to MAR dropout. We show analytically that specific SREMs when fitted to incomplete data, produced by a MAR drop-out mechanism, can yield seriously biased population estimates, with the magnitude of bias increasing as the MAR dropout probability increases. We also propose an alternative SREM model.

In Chapter 4, we explore the consequences of covariance structure misspecification in LMMs under MAR dropout in terms of bias in the population estimates. Specifically, we analytically calculate the asymptotic bias in the marker trend estimates of LMMs

using covariance structures that have been used in real-data applications. We also investigate the performance of LMMs with more complex covariance structures and, in addition, we adopt o Bayesian model comparison to discriminate in real applications among alternative covariance structure specifications of LMMs.

Chapter 5 is concerned with joint modeling of longitudinal data and competing risks. More specifically, we propose a unified and flexible approach to jointly model a normally distributed continuous disease marker (e.g. a transformation of CD4 counts) over time and competing risk data (e.g. death and disengagement from care) using CIFs for the failure submodels, also accounting for failure cause misclassification. Based solely on the assumed joint model, we estimate state occupation and transition probabilities defined jointly by competing risks and discretized marker data. The performance of the proposed methodology is evaluated through simulation studies. We also fit the proposed approaches to CD4 data after ART initiation from the East Africa IeDEA study.

Finally, Chapter 6 first summarizes the relevant literature and discusses the gaps this thesis has filled. In addition, the main developments of this work are briefly presented, discussing also their practical contributions as well as their potential extensions and limitations.

Chapter 2

Overview of longitudinal data analyses methods under missingness

Studies involving repeated measurements of some variable for each individual, or more precisely for each measurement unit (e.g., human beings, animals, or laboratory samples), have become very popular nowadays. Especially in medical research, it is common that a cohort of patients is followed up over time, with plenty of information recorded at a set of pre-specified visits. Thus, a series of longitudinal measurements on the same individual is obtained, with such studies being referred to as longitudinal studies. It has been shown that longitudinal studies concerning a disease play a prominent role in enhancing our understanding of the development and persistence of a disease. The analysis of longitudinal data, though, poses some difficulties compared to other classic approaches in biostatistics, mainly due to the lack of independence of observations from the same individual. The growing popularity of longitudinal studies resulted in the development of numerous statistical approaches that differ in complexity, statistical power, and conditions of use. In longitudinal studies, and more often in studies involving humans, missing data is a frequently encountered problem as it can ultimately

undermine the validity of results under certain conditions.

In this chapter, we first introduce the notation used for longitudinal data with missingness, specifically in the case of monotone missingness in the form of dropout. In Section 2.2, we outline the standard taxonomy of missing data mechanisms used in the literature based on Rubin (1976), also presenting the corresponding implications for data analysis. In Section 2.3, we describe some standard longitudinal analysis approaches that model the observed longitudinal data only. However, these methods are valid only under certain assumptions about the missing data. Finally, in Section 2.4, we introduce the most commonly applied approaches to jointly model longitudinal data and the missing data mechanism, since joint modeling is required in order to obtain reliable estimates under certain assumptions about the missing data mechanism.

2.1 Notation

Let Y_{ij} denote the measurement of the outcome (e.g. a disease marker) intended to be collected at time t_{ij} on a randomly sampled individual i, $j = 1, 2, \ldots, Q_i$, with Q_i denoting the maximum number of marker observations of individual i. In some cases, e.g. in the clinical trials settings, the maximum number of observations and the measurement times are the same for all individuals, i.e. $Q_i = Q$ and $t_{ij} = t_j$, for $i = 1, 2, \ldots, N$, respectively, where N denotes the number of individuals. In other words, all individuals are measured at the same times. These are referred to as balanced data, but this is rarely the case in longitudinal studies including human subjects. Even when the number of marker observations is the same for all individuals, mistimed observations are frequent. Associated with each Y_{ij} are the values x_{ijk}, $k = 1, 2, \ldots, p$, of p explanatory variable that typically include some function of time.

In most cases, though, the data can be incomplete due to some stochastic missing data mechanism; to take missingness into account, let R_{ij} be the missingness indicator for Y_{ij}, i.e. $R_{ij} = 1$ if Y_{ij} is observed and $R_{ij} = 0$ otherwise. In matrix notation, let $\boldsymbol{Y}_i^\top = (Y_{i1}, \ldots, Y_{iQ_i})$ be the full data on individual i and $\boldsymbol{R}_i^\top = (R_{i1}, \ldots, R_{iQ_i})$

the associated missingness indicators. In general, there are 2^{Q_i} possible missing data patterns. To distinguish the data actually observed from the full data, for a given missingness pattern $\boldsymbol{R}_i = \boldsymbol{r}_i$, let $\boldsymbol{Y}_{i,(\boldsymbol{r}_i)}$ denote the observed marker data and $\boldsymbol{Y}_{i,(\overline{\boldsymbol{r}_i})}$ the missing ones. For example, if $Q_i = 3$, given $\boldsymbol{r}_i^\top = (1,0,1)$, $\boldsymbol{Y}_{i,(\boldsymbol{r}_i)} = (Y_{i1}, Y_{i3})^\top$ and $\boldsymbol{Y}_{i,(\overline{\boldsymbol{r}_i})} = Y_{i2}$, that is, the second scheduled measurement is missing. Thus, it is evident that the observed data are represented by $(\boldsymbol{R}_i, \boldsymbol{Y}_{i,(\boldsymbol{R}_i)})$ as both $\boldsymbol{R}_i$ and $\boldsymbol{Y}_{i,(\boldsymbol{R}_i)}$ are required to fully characterize which components of $\boldsymbol{Y}_i$ are observed for a randomly chosen individual. The special focus of this thesis is on monotone missingness mechanisms, e.g. arising from dropout or attrition, which means that the marker data are not available after a certain point in time. This implies that the missing data indicators $\boldsymbol{r}_i$ are of the form $(1,0,\ldots,0)$, $(1,1,\ldots,0)$, …, $(1,1,\ldots,1)$, that is, the subject dropped out after the first, or the second visit, or he/she was followed up until the end of the study. Under this setting, the missing data indicators can be equivalently expressed by the scalar variable $M_i = \sum_{j=1}^{Q_i} R_{ij}$ denoting the number of the observed marker measurements. Thus, given $M_i = m_i$, the observed measurements are $\boldsymbol{Y}_{i,(m_i)}^\top = (Y_{i1}, Y_{i2}, \ldots, Y_{im_i})$ and the missing ones $\boldsymbol{Y}_{i,(\overline{\boldsymbol{m}_i})}^\top = (Y_{im_i+1}, Y_{im_i+2}, \ldots, Y_{iQ_i})$. Most statistical methods aim to provide inferences on the parameters of the distribution of the full marker data $\boldsymbol{Y}_i$, though it is apparent that the estimating process can be based only on the observed marker data $\boldsymbol{Y}_{i,(m_i)}$ of each individual i.

2.2 Inference under missing data

When data are incomplete due to some stochastic mechanism, the density of the full data is equal to

$$f(\boldsymbol{Y}_i, M_i; \boldsymbol{\theta}_L, \boldsymbol{\theta}_t),$$

where $\boldsymbol{\theta}_L$ denotes the parameter vector of the marker model, $f(\boldsymbol{Y}_i; \boldsymbol{\theta}_L)$, and $\boldsymbol{\theta}_t$ denotes the parameter vector of the dropout mechanism, $\Pr(M_i|\boldsymbol{Y}_i; \boldsymbol{\theta}_t)$. Dependence on potential covariates is suppressed here to simplify the notation. It should be noted that it is always valid to factorize $f(\boldsymbol{Y}_i, M_i; \boldsymbol{\theta}_L, \boldsymbol{\theta}_t)$ as $f(\boldsymbol{Y}_i; \boldsymbol{\theta}_L) \times f(M_i|\boldsymbol{Y}_i; \boldsymbol{\theta}_t)$. It has been

shown (Laird, 1988, Rubin, 1976) that when there are missing data, given $M_i = m_i$ for $i = 1, 2, \ldots, N$, the "observed" data likelihood is obtained after integrating the missing data out of the full data density, i.e.

$$\prod_{i=1}^{N} \int f(\boldsymbol{Y}_i, m_i; \boldsymbol{\theta}_L, \boldsymbol{\theta}_t) d\boldsymbol{Y}_{i,(\overline{\boldsymbol{m}_i})} = \prod_{i=1}^{N} \int f(\boldsymbol{Y}_i; \boldsymbol{\theta}_L) \Pr(M_i = m_i | \boldsymbol{Y}_i; \boldsymbol{\theta}_t) d\boldsymbol{Y}_{i,(\overline{\boldsymbol{m}_i})}. \quad (2.1)$$

Therefore, for likelihood-based inferences, we require a parametric model for the joint distribution of $(\boldsymbol{Y}_i, M_i)$. As shown in the literature, the main factor influencing the validity of the methods applied to incomplete data is the form of the dropout mechanism, i.e. $\Pr(M_i | \boldsymbol{Y}_i; \boldsymbol{\theta}_t)$. Standard taxonomy of the missing data mechanisms includes (i) the missing completely at random (MCAR), (ii) the missing at random (MAR), and (iii) the missing not at random (MNAR) definitions, provided below.

Missing completely at random

Data are said to be missing completely at random (MCAR) when the missingness probabilities are unrelated to both the specific marker values that, in principle, should have been obtained and the set of the observed measurements. That is, under monotone missingness, data are MCAR when M_i is independent of both $\boldsymbol{Y}_{i,(M_i)}$ and $\boldsymbol{Y}_{i,(\overline{M_i})}$, the observed and unobserved components of $\boldsymbol{Y}_i$, respectively. As such, missingness in $\boldsymbol{Y}_i$ is simply the result of a chance mechanism unrelated to both observed and unobserved marker measurements. Formally, data are MCAR when

$$\Pr(M_i = m_i | \boldsymbol{Y}_{i,(\boldsymbol{m}_i)}, \boldsymbol{Y}_{i,(\overline{\boldsymbol{m}_i})}; \boldsymbol{\theta}_t) = \Pr(M_i = m_i; \boldsymbol{\theta}_t), \quad (2.2)$$

for each $M_i = 1, 2, \ldots, Q_i$. For example, consider a randomized clinical trial in which participants are randomized at baseline to receive one treatment or the other and then followed for a specified period of time. Subjects have the outcome of interest and other health status information measured at baseline, i.e. immediately prior to starting treatment, and then are to return to the clinic at additional pre-specified times at which the outcome of interest and other health information are to be ascertained. Suppose that individuals drop out of the study because they move away for work or family

considerations; in this case, dropout has nothing to do with the individual's health status or other issues of the study. Thus, it can be reasonable to assume that dropout is MCAR.

Another example of MCAR missingness may come from a nutrition study focusing on the daily average percent fat intake. Accurate measurement of this outcome requires individuals to keep a detailed "food diary" over a long period of time, which is both time-consuming for individuals to maintain and expensive for investigators to design. A simpler approach would be to have individuals recall all the food they ate in the last 24 hours. However, such a measure can only be considered as a surrogate for the outcome of interest as it obviously includes measurement error. To reduce costs and subject burden, the study could be designed so that all individuals provide a 24-hour recall (surrogate) measurement, but only a subset of participants provide the expensive and time-consuming exact measurement, often being referred to as a validation sample. If the validation sample is constructed by selecting members from the entire study sample with probability 0.1, say, then missingness is by design, thus it can be reasonably assumed to be MCAR.

In the statistical literature, it seems that there is no universal agreement on whether the definition of MCAR also assumes no dependence of missingness on the covariates, $\boldsymbol{X}_i$. Little (1995) suggested restricting the use of the term MCAR to the case where $\Pr(M_i = m_i|\boldsymbol{Y}_{i,(\boldsymbol{m}_i)}, \boldsymbol{Y}_{i,(\overline{\boldsymbol{m}_i})}, \boldsymbol{X}_i; \boldsymbol{\theta}_t) = \Pr(M_i = m_i; \boldsymbol{\theta}_t)$, that is, missingness does not depend on observed covariate values. When missingness depends on $\boldsymbol{X}_i$, but is conditionally independent of $\boldsymbol{Y}_i$, $\Pr(M_i = m_i|\boldsymbol{Y}_{i,(\boldsymbol{m}_i)}, \boldsymbol{Y}_{i,(\overline{\boldsymbol{m}_i})}, \boldsymbol{X}_i; \boldsymbol{\theta}_t) = \Pr(M_i = m_i|\boldsymbol{X}_i; \boldsymbol{\theta}_t)$ a subtle, but important, issue arises. The conditional independence of $\boldsymbol{Y}_i$ and $\boldsymbol{R}_i$, given $\boldsymbol{X}_i$, may not hold when conditioning on only a subset of the covariates. Consequently, when an analysis is based on a subset of $\boldsymbol{X}_i$ that excludes a covariate predictive of $\boldsymbol{R}_i$, $\boldsymbol{Y}_i$ is no longer unrelated to $\boldsymbol{R}_i$. To avoid any potential ambiguity, Little (1995) suggests that MCAR be reserved for the case where there is no dependence of $\boldsymbol{R}_i$ on $\boldsymbol{Y}_i$ or $\boldsymbol{X}_i$; when there is dependence on $\boldsymbol{X}_i$ alone, he suggests that the missing data mechanism be referred to as "covariate-dependent" missingness.

The essential feature of MCAR is that the observed data can be thought of as a random sample of the full data. Consequently, all moments, and even the joint distribution, of the observed data do not differ from the corresponding moments or joint distribution of the complete data. Thus, in general, all methods for analysis that yield valid inferences in the absence of missing data will also yield valid inferences when the analysis is based on all available data, or even when it is restricted to the individuals who did not drop out (analysis of completers), although an issue of reduced power could be raised in such an analysis.

Missing at random

In contrast to MCAR, data are said to be missing at random (MAR) when the missingness probabilities depend on the set of observed marker measurements responses, but are further unrelated to the specific missing values that, in principle, should have been obtained. In other words, if individuals are stratified based on similar observed marker values, missingness is just a chance mechanism that does depend on future unobserved values. Formally, dropout is MAR when, given relevant covariates,

$$\Pr(M_i = m_i|\boldsymbol{Y}_{i,(\boldsymbol{m}_i)}, \boldsymbol{Y}_{i,(\overline{\boldsymbol{m}_i})}; \boldsymbol{\theta}_t) = \Pr(M_i = m_i|\boldsymbol{Y}_{i,(\boldsymbol{m}_i)}; \boldsymbol{\theta}_t), \tag{2.3}$$

for each dropout pattern, $j = 1, 2, \ldots, Q_i$, and any realization of the observed data, $\boldsymbol{Y}_{i,(\boldsymbol{m}_i)}$ (Seaman et al., 2013). For a more technical review of MAR definitions, we refer to Seaman et al. (2013). This makes explicit that it is possible to think of MAR as implying that dropout is a sequentially random process; whether or not a subject still in the study drops out at time m_i is "at random" in the sense that it depends on the history of observed data to that point. In the previous surrogate measurement example, if the validation sample is constructed by selecting individuals from the full study sample depending on whether or not their 24-hour recall measurements exceed some threshold, then the missingness of the expensive diary measurement depends on a variable that is always observed, and hence is MAR. In the clinical trial example, if drop-out depends on observed health history, then missingness is MAR. If, for example,

all patients who reach a certain value of the disease marker stop trial, then this, most likely, represents a MAR mechanism.

Recall that the likelihood for the observed data is equal to

$$\begin{aligned} & \prod_{i=1}^{N} \int f(\boldsymbol{Y}_i, m_i; \boldsymbol{\theta}_L, \boldsymbol{\theta}_t) d\boldsymbol{Y}_{i,(\overline{\boldsymbol{m}_i})} \\ = & \prod_{i=1}^{N} \int f(\boldsymbol{Y}_i; \boldsymbol{\theta}_L) \Pr(M_i = m_i | \boldsymbol{Y}_{i,(m_i)}, \boldsymbol{Y}_{i,(\overline{\boldsymbol{m}_i})}; \boldsymbol{\theta}_t) d\boldsymbol{Y}_{i,(\overline{\boldsymbol{m}_i})}, \end{aligned} \tag{2.4}$$

that is, the full data density with the missing data integrated out. Under MAR dropout though, $\Pr(M_i = m_i | \boldsymbol{Y}_{i,(m_i)}, \boldsymbol{Y}_{i,(\overline{\boldsymbol{m}_i})}; \boldsymbol{\theta}_t) = \Pr(M_i = m_i | \boldsymbol{Y}_{i,(m_i)}; \boldsymbol{\theta}_t)$ for any value of the missing data, $\boldsymbol{Y}_{i,(\overline{\boldsymbol{m}_i})}$. Thus, Equation (2.4) is equal to

$$\prod_{i=1}^{N} \Pr(M_i = m_i | \boldsymbol{Y}_{i,(m_i)}; \boldsymbol{\theta}_t) \int f(\boldsymbol{Y}_i; \boldsymbol{\theta}_L) d\boldsymbol{Y}_{i,(\overline{\boldsymbol{m}_i})}, \tag{2.5}$$

where $f(\boldsymbol{Y}_{i,(m_i)}; \boldsymbol{\theta}_L) = \int f(\boldsymbol{Y}_i; \boldsymbol{\theta}_L) d\boldsymbol{Y}_{i,(\overline{\boldsymbol{m}_i})} = \int f(\boldsymbol{Y}_{i,(m_i)}, \boldsymbol{Y}_{i,(\overline{\boldsymbol{m}_i})}; \boldsymbol{\theta}_L) d\boldsymbol{Y}_{i,(\overline{\boldsymbol{m}_i})}$, i.e. the density function for the part of $\boldsymbol{Y}_i$ that is observed for fixed $M_i = m_i$, as it is simply the density function for the full marker data with the "missing part" integrated out. For example, if $\boldsymbol{Y}_i \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then the marginal distribution of $\boldsymbol{Y}_{i,(m_i)} \sim N(\boldsymbol{\mu}_{(m_i)}, \boldsymbol{\Sigma}_{(m_i)})$, by standard properties of the multivariate Normal distribution, where $\boldsymbol{\mu}_{(m_i)}$ and $\boldsymbol{\Sigma}_{(m_i)}$ denote the corresponding submatrices of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively. For other multivariate distributions, though, such a property may not hold. If we further assume that $\boldsymbol{\theta}_L$ and $\boldsymbol{\theta}_t$ are variation independent, or "distinct" (Rubin, 1976), i.e. denoting the parameter spaces of $\boldsymbol{\theta}_L$ and $\boldsymbol{\theta}_t$ by $\boldsymbol{\Theta}_L$ and $\boldsymbol{\Theta}_t$, respectively, the parameter space for $(\boldsymbol{\theta}_L^\top, \boldsymbol{\theta}_t^\top)^\top$ is $\boldsymbol{\Theta}_L \times \boldsymbol{\Theta}_t$, it is evident that all information about $\boldsymbol{\theta}_L$ is included in $\prod_{i=1}^{N} \int f(\boldsymbol{Y}_{i,(m_i)}; \boldsymbol{\theta}_L)$ and thus $\prod_{i=1}^{N} \Pr(M_i = m_i | \boldsymbol{Y}_{i,(m_i)}; \boldsymbol{\theta}_t)$ is irrelevant for estimating $\boldsymbol{\theta}_L$. Hence, a MAR dropout mechanism is said to be ignorable in the literature (Laird, 1988, Little and Rubin, 2002, Rubin, 1976), as there is no need to estimate the dropout mechanism when likelihood-based methods are applied for the marker model. That is, the maximum likelihood estimator based on $\prod_{i=1}^{N} \int f(\boldsymbol{Y}_{i,(m_i)}; \boldsymbol{\theta}_L)$ is consistent for $\boldsymbol{\theta}_L$, irrespective of the dropout mechanism (Rubin, 1976). The last statement is critical as an ignorable mechanism implies nothing about the validity of inferences which are not based on the

likelihood of the observed marker data. It should be also noted that it is the missingness mechanism, not the missing data (or the individuals with missing data), which can be ignored (Laird, 1988) as the distribution of $\boldsymbol{Y}_{i,(m_i)}|M_i = m_i$ is no longer equal to the marginal distribution of $\boldsymbol{Y}_{i,(m_i)}$ under MAR. The important implication of this feature of MAR is that sample moments are not unbiased estimates of the same moments of the full data. Thus, inference based on these moments without accounting for MAR, such as scatterplots of the sample marker evolution, may prove misleading.

An equivalent formulation of MAR is in terms of the predictive distribution of the missing marker measurements, $\boldsymbol{Y}_{i,(\overline{\boldsymbol{m}_i})}$, given the observed ones, $\boldsymbol{Y}_{i,(m_i)}$ and $M_i = m_i$, has been provided in Molenberghs et al. (2008). In particular, assuming MAR, this distribution takes the following form:

$$f(\boldsymbol{Y}_{i,(\overline{\boldsymbol{m}_i})}|\boldsymbol{Y}_{i,(m_i)}, M_i = m_i; \boldsymbol{\theta}_L, \boldsymbol{\theta}_t) = f(\boldsymbol{Y}_{i,(\overline{\boldsymbol{m}_i})}|\boldsymbol{Y}_{i,(m_i)}; \boldsymbol{\theta}_L), \tag{2.6}$$

i.e. only the distribution of each individual's missing values $\boldsymbol{Y}_{i,(\overline{\boldsymbol{m}_i})}$, conditioned on the observed values, $\boldsymbol{Y}_{i,(m_i)}$, is the same as the distribution of the corresponding observations of the full data. Thus, missing values can be validly predicted using the observed data and a correctly specified model for the joint distribution of $(\boldsymbol{Y}^{\top}_{i,(m_i)}, \boldsymbol{Y}^{\top}_{i,(\overline{\boldsymbol{m}_i})})$.

Missing not at random

The third type of missing data mechanisms is referred to as missing not at random (MNAR). In contrast to MAR, missing data are said to be MNAR when the probability that marker measurements are missing is related to the specific values that should have been obtained, in addition to the ones actually observed. That is, in the case of dropout, the conditional distribution of M_i, given $\boldsymbol{Y}_{i,(m_i)}$, is related to $\boldsymbol{Y}_{i,(\overline{\boldsymbol{m}_i})}$, with $\Pr(M_i|\boldsymbol{Y}_i, \boldsymbol{\theta}_t) = \Pr(M_i|\boldsymbol{Y}_{i,(m_i)}, \boldsymbol{Y}_{i,(\overline{\boldsymbol{m}_i})}; \boldsymbol{\theta}_t)$ depending on at least ons component of $\boldsymbol{Y}_{i,(\overline{\boldsymbol{m}_i})}$.

For example, consider a longitudinal study comparing anti-hypertensive agents with systolic blood pressure as the outcome. Suppose that the outcome is to be collected at pre-specified scheduled visit times. An individual might take his/her blood pressure

right before the scheduled clinic visit and, if it is high, decide to terminate his/her participation in the study, feeling the agent he/she is taking is not helping. This a classic example of MNAR dropout, where at time t_{ij}, the probability of dropout increases with the (unobserved) outcome value, Y_{ij}.

An MNAR mechanism is often referred to as non-ignorable missingness because the missing data mechanism cannot be ignored when one aims to make inferences about the distribution of the complete marker data. This is the case as the integral in Equation (2.4) does no longer simplify to Equation (2.5). In general, any valid inferential procedure under MNAR requires specification of a model for the missing data mechanism. We note that the term informative has often been used to describe missing data that are MNAR, especially for the monotone missingness such as dropout.

When dropout is MNAR, the model assumed for $\Pr(M_i|\boldsymbol{Y}_{i,(m_i)}, \boldsymbol{Y}_{i,(\overline{\boldsymbol{m}_i})}; \boldsymbol{\theta}_t)$ is critical and must be included in the analysis. Moreover, many authors have pointed out that the results are often sensitive to the specified missingness model, which can fully drive the results in some cases. It has also been shown that an MNAR assumption about the missingness mechanism cannot be ultimately verified based solely on the observed data (Molenberghs et al., 2008). That is, the observed data provide no information that can either fully support or refute one MNAR mechanism over another. Thus, without additional information, it is evident that identification is driven by model assumptions, and many authors have discussed the importance of conducting sensitivity analyses (e.g. Copas and Eguchi, 2001, Creemers et al., 2011, Laird, 1988, Little and Rubin, 2002, Molenberghs et al., 2001, Njagi et al., 2014, Rosenbaum and Rubin, 1983, 1985, Scharfstein et al., 1999, Verbeke et al., 2001). It is, therefore, suggested that, when missingness is thought to be MNAR, the sensitivity of inferences to a variety of plausible assumptions concerning the missingness process should be carefully assessed.

Under MNAR, almost all standard methods of analyzing the observed marker data only (e.g. likelihood-based methods) are invalid. To obtain valid estimators, joint models for the marker measurements and the missing data mechanism are required. Models for $(\boldsymbol{Y}_i, M_i)$ can be specified in numerous ways, depending on how the joint

distribution is factorized. Three common factorizations of the joint distribution of the marker and missing data mechanism give rise to selection, pattern-mixture, and shared-parameter models, which are described in Section 2.4.

We first review in Section 2.3 some standard approaches for longitudinal data analysis that model the observed marker data only, and essentially assume that missingness is MCAR or MAR, whereas in 2.4 we review some MNAR approaches which jointly model the observed marker measurements and the dropout probabilities.

2.3 Longitudinal data analysis methods assuming MCAR or MAR

2.3.1 General linear model with correlated errors

As previously mentioned, when data are incomplete due to some stochastic missing data mechanism, most statistical methods assume a model for the full data intended to be collected. When the outcomes Y_{ij}, $j = 1, 2, \ldots, Q_i$, are continuous, one traditional approach is the general linear model, which assumes that Y_{ij} follow the regression model:

$$Y_{ij} = \beta_1 x_{ij1} + \ldots + \beta_p x_{ijp} + \epsilon_{ij},$$

where β_j, $j = 1, 2, \ldots, p$ are the regression coefficients (or the fixed effects, which are constant across all individuals) and ϵ_{ij} are error terms assumed to have zero mean which, on the contrary to the classical linear model, are assumed to be correlated as they refer to the same individual. Thus, apart from modeling $E(Y_{ij})$ over time, longitudinal data analysis requires modeling the covariance structure of the data, $Cov(Y_{ij}, Y_{ik})$, $j \neq k$, as well. In matrix notation, recall that $\boldsymbol{Y}_i^\top = (Y_1, \ldots, Y_{Q_i})$ is the full set of marker measurements, $\boldsymbol{X}_i$ the $Q_i \times p$ fixed effects design matrix with the jth row equal to $(x_{ij1}, \ldots, x_{ijp})$, and $\boldsymbol{\epsilon}_i^\top = (\epsilon_{i1}, \ldots, \epsilon_{iQ_i})$ the full vector of errors. To explicitly model the correlation in the data, the covariance matrix of $\boldsymbol{\epsilon}_i$ is assumed to be equal to $Var(\boldsymbol{\epsilon}_i) = \sigma^2 \boldsymbol{V}_i$, where $\boldsymbol{V}_i$ is a non-diagonal positive definite matrix, which further

implies that $E(\boldsymbol{Y}_i|\boldsymbol{X}_i) = \boldsymbol{X}_i\boldsymbol{\beta}$ and $Var(\boldsymbol{Y}_i|\boldsymbol{X}_i) = \sigma^2\boldsymbol{V}_i$. The form of $\boldsymbol{V}_i$ is crucial for robust inference as reliable inference on the fixed effects (which are of primary interest) requires that the form of $\boldsymbol{V}_i$ can adequately describe the true covariance structure of the data. Typically, the form of $\boldsymbol{V}_i$ is expected to be a function of time, depending on some unknown parameters ($\boldsymbol{\theta}_v$) to be estimated based on the data. Some traditional approaches for modeling the covariance are described below (e.g. Diggle et al., 2002, Fitzmaurice et al., 2008, 2011).

Compound symmetry

Historically, one of the first covariance pattern models used for analysis of repeated measurements was the compound symmetry. With a compound symmetry covariance, it is assumed that the variance is constant across occasions, say σ^2 , and $Cor(Y_{ij}, Y_{ik}) = \rho$ for all j and k. That is,

$$Var(\boldsymbol{Y}_i) = \sigma^2 \begin{pmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \rho & \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \cdots & 1 \end{pmatrix},$$

usually with the constraint that $\rho > 0$. The compound symmetry covariance is very parsimonious, with only two parameters regardless of the number of visits/occasions and whether the data are balanced or not. Thus, it is mainly applied for explanatory analyses due to its simplicity. Also, it has a randomization justification in certain settings (e.g., split-plot designs) and it also equivalent to a random-effects model with a single random effect per individual. However, it does make the rather strong assumption that the correlation between any pair of measurements is the same regardless of the time interval between the measurements. This latter aspect of the compound symmetry covariance, i.e. the constraint on the correlation among repeated measurements, is somewhat unappealing for most longitudinal data, where the correlations are expected

to decay over time. Also the assumption of constant variance across time may be unrealistic in most settings.

Exponential correlation

The motivation for the Exponential correlation model derives from the fact that the correlation of observations closer together in time is expected to be higher than that of observations farther apart. Formally, the correlation between two marker measurements is assumed to depend on the time distance between two measurements, i.e.

$$Cor(Y_{ij}, Y_{ik}) = \rho^{|t_{ij}-t_{ik}|},$$

with the variance assumed to be constant across all measurement occasions, σ^2. That is, the correlation between any pair of repeated measures decreases exponentially with the time separations between them. This structure is referred to as an "exponential" covariance model because it can be re-expressed as

$$\begin{aligned} Cov(Y_{ij}, Y_{ik}) &= \sigma^2 \rho^{|t_{ij}-t_{ik}|}, \\ &= \sigma^2 \exp\left(-\iota |t_{ij} - t_{ik}|\right), \end{aligned}$$

where $\iota = -\log(\rho)$ for $\iota > 0$. Also note that the Exponential covariance model is invariant under linear transformations of the time scale. That is, if we replace t_{ij} by $(a+bt_{ij})$ (e.g., if we replace "weeks" by "days"), the same form for the covariance matrix holds. The main advantage of the Exponential correlation is that it is much more realistic than the compound symmetry model, as the correlation is time-dependent, while remaining parsimonious as it includes only two parameters. Thus, it can be applied to real data irrespective of whether the data are balanced or not. A distinctive feature of the Exponential correlation model is that it assumes that the correlation is equal to one if measurements are made repeatedly at the same occasion (or replicate measurements on an individual can be obtained at the same occasion), and that the correlation decreases rapidly to zero as the time separation between measurements increases. This first aspect of the exponential covariance model corresponds to the

assumption that the responses are measured without error, an unrealistic assumption in most longitudinal studies in health sciences. The latter feature, correlations among repeated measurements that decay to zero, is rarely observed in longitudinal studies.

Gaussian correlation

In the Exponential covariance model, the log correlation is linear in the distance. An alternative covariance model is the Gaussian correlation model, which specifies the decay of correlation using the squared distance:

$$Cor(Y_{ij}, Y_{ik}) = \exp\left\{-\iota(t_{ij} - t_{ik})^2\right\},$$

where $\iota > 0$. The Gaussian correlation model is very similar to the Exponential correlation model, so the choice between one of them should be based on model fit to the observed data.

Unstructured covariance

When the number of measurement occasions is relatively small and the (observed) marker data are roughly balanced, it may be reasonable to allow the covariance matrix to be arbitrary, with all of its elements unconstrained. The only formal requirement, though, is that the covariance matrix be symmetric and positive-definite. When no explicit structure is assumed for the covariance among the repeated measures, the resulting covariance is referred to as an "unstructured" covariance. The main advantage of such an "unstructured" covariance is that no assumptions are made about the variances and covariances. The absence of restrictions on the variances is essential since empirical evidence based on various longitudinal studies suggests that the variances are rarely constant over time. With n measurement occasions, the "unstructured" covariance matrix has $n \times (n+1)/2$ parameters: the n variances and then $n \times (n-1)/2$ pairwise covariances (or correlations).

The main drawback, though, is that the number of covariance parameters to be estimated grows rapidly with the number of occasions. For example, when there are

three occasions ($n = 3$), the number of covariance parameters is 6 (3 variances and 3 pairwise covariances). However, when $n = 5$, the number of covariance parameters has grown to 15, while when $n = 10$, the number of covariance parameters is 55. When the number of covariance parameters that need to be estimated is large, relative to the sample size, estimation is likely to be highly unstable. Thus, using an unstructured covariance will be appealing only in cases where the number of subjects, N, is large relative to the number of covariance parameters, $n \times (n + 1)/2$. Some methods that improve the convergence of the unstructured covariance model have been proposed in the literature (Lu and Mehrotra, 2010). Also, under certain assumptions, it has been shown that the unstructured covariance model can be marginally equivalent to a mixed model using B-splines for the random effects, as explained in detail in García-Hernandez et al. (2020).

Setting aside the issue of the potentially large number of covariance parameters that may need to be estimated, the use of an unstructured covariance matrix is problematic when there are mistimed measurements or, more generally, when visiting times are expected to be irregular. Even the most carefully designed longitudinal study will frequently suffer from deviations from the measurement protocol, resulting in measurements made at arbitrary, irregularly timed intervals. When this problem arises, as it frequently does in studies in health sciences, the resulting mistimed repeated measurements cannot be accommodated in an unstructured covariance. Thus, when the longitudinal data are inherently unbalanced and/or when the sample size is not sufficiently large to estimate an unstructured covariance, it is usually desirable to impose some structure on the covariance matrix.

Estimating methods

Since the desired complete data are not available, statistical procedures to estimate the model parameters can take into account only the data that are actually observed. A traditional approach to semi-parametrically estimate the fixed effects, $\boldsymbol{\beta}$, is to use the weighted least squares estimator, $\hat{\boldsymbol{\beta}}(\boldsymbol{W})$, which minimizes the sum of the quadratic

forms $(\boldsymbol{Y}_{i,(m_i)} - \boldsymbol{X}_{i,(m_i)}\boldsymbol{\beta})^\top \boldsymbol{W}_{i,(m_i)}(\boldsymbol{Y} - \boldsymbol{X}_{i,(m_i)}\boldsymbol{\beta})$ over subjects, where $\boldsymbol{W}_{i,(m_i)}$ is any symmetric "weight" matrix. Using straightforward manipulations, it can be shown that the weighted least squares estimator, conditionally on observing the dropout pattern indicators, $M_i = m_i, i = 1, 2, \ldots, N$, is equal to

$$\hat{\boldsymbol{\beta}}(\boldsymbol{W}) = \left(\sum_{i=1}^{N} \boldsymbol{X}_{i,(m_i)}^\top \boldsymbol{W}_{i,(m_i)} \boldsymbol{X}_{i,(m_i)}\right)^{-1} \sum_{i=1}^{N} \boldsymbol{X}_{i,(m_i)}^\top \boldsymbol{W}_{i,(m_i)} \boldsymbol{Y}_{i,(m_i)},$$

while it is straightforward to verify that $\hat{\boldsymbol{\beta}}(\boldsymbol{W})$ is unbiased for $\boldsymbol{\beta}$, i.e. $E\{\hat{\boldsymbol{\beta}}(\boldsymbol{W})\} = \boldsymbol{\beta}$, irrespective of the choice for $\boldsymbol{W}_{i,(m_i)}$ if $E(\boldsymbol{Y}_{i,(m_i)}|M_i = m_i) = \boldsymbol{X}_{i,(m_i)}\boldsymbol{\beta}$. This happens if either (i) there are no missing data at all (i.e. $M_i = Q_i$, for all $i = 1, 2, \ldots, N$) or (ii) $\boldsymbol{Y}_i$ and M_i are independent random variables, which is equivalent to an MCAR assumption for dropout. The corresponding variance, though, is equal to

$$Var\left\{\hat{\boldsymbol{\beta}}(\boldsymbol{W})\right\} = \sigma^2 \left(\sum_{i=1}^{N} \boldsymbol{X}_{i,(m_i)}^\top \boldsymbol{W}_{i,(m_i)} \boldsymbol{X}_{i,(m_i)}\right)^{-1}$$
$$\left(\sum_{i=1}^{N} \boldsymbol{X}_{i,(m_i)}^\top \boldsymbol{W}_{i,(m_i)} \boldsymbol{V}_{i,(m_i)} \boldsymbol{W}_{i,(m_i)} \boldsymbol{X}_{i,(m_i)}\right) \left(\sum_{i=1}^{N} \boldsymbol{X}_{i,(m_i)}^\top \boldsymbol{W}_{i,(m_i)} \boldsymbol{X}_{i,(m_i)}\right)^{-1},$$

which depends on both $\boldsymbol{W}_{i,(m_i)}$ and $\boldsymbol{V}_{i,(m_i)}$. This last remark suggests that the most efficient weighted least-squares estimator for $\boldsymbol{\beta}$ uses $\boldsymbol{W}_{i,(m_i)} = \boldsymbol{V}_{i,(m_i)}^{-1}$ (Diggle et al., 2002), in which case $Var\left\{\hat{\boldsymbol{\beta}}(\boldsymbol{V})\right\} = \sigma^2 \left(\sum_{i=1}^{N} \boldsymbol{X}_{i,(m_i)}^\top \boldsymbol{V}_{i,(m_i)}^{-1} \boldsymbol{X}_{i,(m_i)}\right)^{-1}$. Thus, it is evident that the specification of the variance matrix, $\boldsymbol{V}_{i,(m_i)}$, is crucial when modeling longitudinal data even though any weight matrix can produce unbiased estimates under MCAR or no missing data at all. Fully semi-parametric estimation of $\boldsymbol{\beta}(\boldsymbol{V})$ and $\boldsymbol{\theta}_v$ could be carried out using the generalized estimating equation (GEE) theory described in Section 2.3.2.

An alternative estimation procedure is to assume a fully parametric model based on the Normal distribution for the full data, $\boldsymbol{Y}_i \sim N(\boldsymbol{X}_i\boldsymbol{\beta}, \sigma^2\boldsymbol{V}_i)$. Thus, the marginal distribution of the observed data is equal to $\boldsymbol{Y}_{i,(m_i)} \sim N(\boldsymbol{X}_{i,(m_i)}\boldsymbol{\beta}, \sigma^2\boldsymbol{V}_{i,(m_i)})$, with parameter estimates obtained by maximizing the likelihood function of the observed

data (conditionally on $M_i = m_i, i = 1, 2, \ldots, N$):

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\theta}_v) &= -\frac{1}{2}\Bigg\{ n \log(\sigma^2) + \sum_{i=1}^{N} \log |\boldsymbol{V}_{i,(m_i)}| \\ &+ \frac{1}{\sigma^2} \sum_{i=1}^{N} (\boldsymbol{Y}_{i,(m_i)} - \boldsymbol{X}_{i,(m_i)}\boldsymbol{\beta})^\top \boldsymbol{V}_{i,(m_i)}^{-1} (\boldsymbol{Y}_{i,(m_i)} - \boldsymbol{X}_{i,(m_i)}\boldsymbol{\beta}) \Bigg\}, \quad (2.7) \end{aligned}$$

where $n = \sum_{i=1}^{N} m_i$. Given $\boldsymbol{V}_{i,(m_i)}$, $i = 1, 2, \ldots, N$, the maximum likelihood estimator for $\boldsymbol{\beta}$ coincides with the weighted least squared estimator

$$\hat{\boldsymbol{\beta}}(\boldsymbol{V}) = \left(\sum_{i=1}^{N} \boldsymbol{X}_{i,(m_i)}^\top \boldsymbol{V}_{i,(m_i)}^{-1} \boldsymbol{X}_{i,(m_i)} \right)^{-1} \sum_{i=1}^{N} \boldsymbol{X}_{i,(m_i)}^\top \boldsymbol{V}_{i,(m_i)}^{-1} \boldsymbol{Y}_{i,(m_i)},$$

replacing $\boldsymbol{W}_{i,(m_i)}$ by $\boldsymbol{V}_{i,(m_i)}^{-1}$. The maximum likelihood estimate for σ^2 (given $\boldsymbol{V}$) is equal to

$$\hat{\sigma^2}(\boldsymbol{V}) = \sum_{i=1}^{N} \{\boldsymbol{Y}_{i,(m_i)} - \boldsymbol{X}_{i,(m_i)}\hat{\boldsymbol{\beta}}(\boldsymbol{V})\}^\top \boldsymbol{V}_{i,(m_i)}^{-1} \{\boldsymbol{Y}_{i,(m_i)} - \boldsymbol{X}_{i,(m_i)}\hat{\boldsymbol{\beta}}(\boldsymbol{V})\} \Big/ n,$$

whereas maximization with respect to the unique parameters in $\boldsymbol{V}$ can be carried out after substituting $\boldsymbol{\beta} \equiv \hat{\boldsymbol{\beta}}(\boldsymbol{V})$ and $\sigma^2 \equiv \hat{\sigma^2}(\boldsymbol{V})$ into the likelihood function (2.7). Except for very special cases, though, this requires a numerical optimization routine, such as the Newton-Raphson algorithm.

Note that when using maximum likelihood estimation for simultaneous estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}_v$, the design matrix $\boldsymbol{X}_{i,(m_i)}$ is involved explicitly in the estimation of $\boldsymbol{\theta}_v$. One consequence of this is that if assuming an incorrect form for $\boldsymbol{X}_{i,(m_i)}$, we may get inconsistent estimates for $\boldsymbol{\theta}_v$. To address this issue, it is recommended (Diggle et al., 2002) to use an over-elaborated form for $\boldsymbol{X}_{i,(m_i)}$ when investigating the covariance structure of the data. When the data come from a well-designed experiment, it is suggested to use a "saturated" mean model by using time as a class factor for each occasion. After inferring the most appropriate covariance model, a more parsimonious mean model can be used. However, in most observational studies, there are continuous covariates and the visiting times may be irregular, so the concept of a saturated model may break down.

Another potential problem with the saturated model is that when the number of columns in $\boldsymbol{X}_{i,(m_i)}$ is large relatively to the number of individuals, the estimates for $\boldsymbol{\theta}_v$ can be seriously biased. For example, it is well known that when $\boldsymbol{V}_{i,(m_i)} = \boldsymbol{I}_{(m_i)}$, an unbiased estimator for σ^2 requires a divisor of $(n-p)$, rather than n, in the denominator, and the problem is further exacerbated by the correlated structure of the data. To overcome this issue, one can use the method of restricted maximum likelihood (REML) (Patterson and Thompson, 1971). The REML estimator is defined as a maximum likelihood estimator based on a linearly transformed set of data such that the resulting distribution does not depend on $\boldsymbol{\beta}$. One way to achieve this is by using the projection matrix converting the data to ordinary least squares residuals. Some main findings include:

- The resulted estimators for σ^2 and $\boldsymbol{V}$ do not depend on the choice of the transformation matrix.
- The transformation needs not be explicit.

After noting that the Jacobian of the transformation does not depend on the model parameters, it can be shown that the REML is equal to Equation (2.7) plus the term $-\frac{1}{2}\log|\sigma^{-2}\sum_{i=1}^{N}\boldsymbol{X}_{i,(m_i)}^{\top}\boldsymbol{V}_{i,(m_i)}^{-1}\boldsymbol{X}_{i,(m_i)}|$. Therefore, given $\boldsymbol{V}_{i,(m_i)}$, the REML estimates for $\boldsymbol{\beta}$ and σ are the same, except for $n-p$ instead of n in the denomination of the estimator for σ^2. In general, the REML estimator is asymptotically equivalent to the MLE when p is fixed and $n \to \infty$. It should be noted that since the REML and MLE estimates of the variance components differ, the estimates for $\boldsymbol{\beta}$ also differ, but often not substantially. The REML approach has also a Bayesian interpretation, as it corresponds to using a uniform prior distribution for $\boldsymbol{\beta}$ and integrating it out of the joint posterior distribution.

2.3.2 Population-averaged (marginal) models using GEE

The generalized estimating equations (GEE) approach has been widely used for longitudinal data analysis, especially when the marker is discrete, in which case there is no

well-defined and easy to apply multivariate distribution. GEEs are often also referred to as marginal models, since they model the marginal expected values of the marker for each time point conditionally on explanatory variables, similarly to what is modeled in a cross-sectional study. The popularity of GEEs is mainly due to the lack of full distributional assumptions for the full vector of marker values, i.e. they are semi-parametric, actually extending the idea of quasi-likelihood functions (Wedderburn, 1974). The GEE approach can be though of as being a particular type of extension of generalized linear models (GLM) to longitudinal, or more generally, cluster-correlated data, though GLMs can be also adapted to the longitudinal case by incorporating individual-specific random effects, as well. In GEEs (Liang and Zeger, 1986, Zeger and Liang, 1986, Zeger et al., 1988), the expected value of the marker at visit j, $j = 1, 2, \ldots, Q_i$, $E(Y_{ij}|\boldsymbol{X}_{ij}) = \mu_{ij}$, is assumed to depend on covariates $\boldsymbol{X}_{ij}$, with a known link function $g(\mu_{ij}) = \boldsymbol{X}_{ij}\boldsymbol{\beta}$, with $\boldsymbol{\beta}$, the population-average parameter, being of main interest. The variance of the marker at visit j is assumed to be $Var(Y_{ij}|\boldsymbol{X}_{ij}) = \phi V(\mu_{ij})$, where $V(\cdot)$ is a known variance function and ϕ is a nuisance scale factor. Thus, the mean model follows the GLM structure, whereas dependence between observations is considered as a nuisance parameter in the residual error component of the model. In matrix notation, the expected value is equal to $E(\boldsymbol{Y}_i|\boldsymbol{X}_i) = \boldsymbol{\mu}_i$, with $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{iQ_i})^\top$, and the assumed covariance matrix of $\boldsymbol{Y}_i$, $Cov(\boldsymbol{Y}_i) = \boldsymbol{V}_i = \phi \boldsymbol{C}_i^{1/2} \boldsymbol{R}_i(\rho) \boldsymbol{C}_i^{1/2}$, where $\boldsymbol{R}_i(\rho)$ is an assumed (working) correlation matrix depending on a vector of unknown parameters, ρ, and $\boldsymbol{C}_i = \text{Diag}\{V(\mu_{i1}), \ldots, V(\mu_{iQ_i})\}$. Then the regression parameters $\boldsymbol{\beta}$ are estimated by the GEE

$$\mathcal{U}(\boldsymbol{\beta}) = \sum_{i=1}^{N} \boldsymbol{D}_i^\top \boldsymbol{V}_i^{-1} (\boldsymbol{Y}_i - \boldsymbol{\mu}_i) = \boldsymbol{0}, \tag{2.8}$$

with $\boldsymbol{D}_i = \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}$. However, in most applications, ϕ and ρ would be unknown. Zeger and Liang (1986) proposed the following iterative procedure to obtain estimates for $\boldsymbol{\beta}$: (i) start with an initial estimate of $\boldsymbol{\beta}$, e.g. obtained through ordinary GLM assuming independent observations, (ii) given the current estimate $\boldsymbol{\beta}^{(l)}$, calculate the method-of-moments estimates for ϕ and ρ and estimate the covariance matrix $\hat{\boldsymbol{V}}_i =$

$\hat{\phi}\hat{\boldsymbol{C}}_i^{1/2}\boldsymbol{R}_i(\hat{\rho})\hat{\boldsymbol{C}}_i^{1/2}$, and (iii) update $\boldsymbol{\beta}$ through a modified Fisher scoring algorithm $\boldsymbol{\beta}^{(l+1)} = \boldsymbol{\beta}^{(l)} + \left(\sum_{i=1}^N \hat{\boldsymbol{D}}_i^\top \hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{D}}_i\right)^{-1}\sum_{i=1}^N \hat{\boldsymbol{D}}_i^\top \hat{\boldsymbol{V}}_i^{-1}(\boldsymbol{Y}_i - \hat{\boldsymbol{\mu}}_i)$, iterating between (ii) and (iii) until convergence. In a seminal paper, Liang and Zeger (1986) proved that the estimator $\hat{\boldsymbol{\beta}}$ is consistent and asymptotically Normal irrespective of whether $\boldsymbol{R}_i(\rho)$ is correctly specified, hence the term *working* correlation matrix for $\boldsymbol{R}_i(\rho)$. However, if $\boldsymbol{R}_i(\rho)$ coincides with the true correlation structure of the data, the efficiency of the estimator obtained by solving Equation (2.8) improves. Furthermore, Liang and Zeger (1986) showed that if the correlation matrix is correctly specified, the covariance matrix of $\hat{\boldsymbol{\beta}}$ can be accurately estimated by $\boldsymbol{A}^{-1}$, where $\boldsymbol{A} = \sum_{i=1}^N \hat{\boldsymbol{D}}_i^\top \hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{D}}_i$, whereas if the correlation is not correct, the covariance matrix should be estimated by a sandwich formula, i.e. $\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}$, where $\boldsymbol{B} = \sum_{i=1}^N \hat{\boldsymbol{D}}_i^\top \hat{\boldsymbol{V}}_i^{-1}(\boldsymbol{Y}_i - \hat{\boldsymbol{\mu}}_i)(\boldsymbol{Y}_i - \hat{\boldsymbol{\mu}}_i)^\top \hat{\boldsymbol{V}}_i^{-1}\hat{\boldsymbol{D}}_i$.

Despite the great flexibility of the GEE estimators, their appealing properties may be compromised when there is dropout. Missing data create no difficulties in obtaining GEE estimates as one can circumvent the problem by simply basing inferences on the observed responses, i.e. by solving

$$\mathcal{U}^{obs}(\boldsymbol{\beta}) = \sum_{i=1}^{N} \boldsymbol{D}_{i,(m_i)}^\top \boldsymbol{V}_{i,(m_i)}^{-1}(\boldsymbol{Y}_{i,(m_i)} - \boldsymbol{\mu}_{i,(m_i)}) = \boldsymbol{0}. \tag{2.9}$$

However, consistency of $\hat{\boldsymbol{\beta}}$ when the correlation model is misspecified can be directly verified only when dropout is MCAR (Liang and Zeger, 1986). Consequently, several authors have pointed out when data are MAR, the standard GEE method that bases inferences on the observed marker values can yield biased regression parameter estimates, though this may be not the case if the the correlation model is correctly specified in specific examples (Liang and Zeger, 1986). Therefore, there have been various extensions of the standard GEE method proposed by Liang and Zeger (1986), mainly to improve the estimation of the correlation model (e.g. Liang et al., 1992, Lipsitz et al., 1991, 1992, 2000, Prentice, 1988, Zhao and Prentice, 1990).

When missing data are assumed to be MAR, two main approaches have been proposed to handle this problem within the GEE framework: multiple imputation (e.g. Paik, 1997) and weighted estimating equations (Robins et al., 1995). The idea behind

multiple imputation is very simple: one just needs to fill in the missing marker values with imputed values. However, to properly reflect the uncertainty inherent in the imputation of the unobserved values, the imputation process is repeated multiple times. The imputations are typically based on some assumed model for the missing data given the observed data and other relevant covariates. The nice feature of imputation methods is that, once a filled-in data set has been constructed, the standard GEE method for complete data can be readily applied. The multiple data sets yield different sets of parameter estimates and standard errors, which are appropriately combined to provide a single parameter estimate together with a standard error obtained from Rubin's rule (Rubin, 1987). In any case, multiple imputation methods have been very popular in recent years as they can be straightforwardly applied to any model when missingness arises.

The second general approach for handling data that are MAR is via weighted estimating equations, where an individual's contribution to the standard GEE is weighted inversely by the probability of being observed at the given times. Thus, the fact that certain marker profiles are under-represented in the observed cohort is formally taken into account. The idea originates from the survey literature (Horvitz and Thompson, 1952) in which the weights are often known and related to the survey design. To clarify the weighting methods, we present a simple example where there are at most two marker measurements and the first one is always observed. Then the probability of a complete case is assumed to be equal to $\pi_i = \pi_i(\boldsymbol{\theta}_t) = \Pr(M_i = 2|Y_{i1}; \boldsymbol{\theta}_t)$, i.e. MAR, since the probability of dropout depends on the first marker measurement, which is always observed. Note also that $\Pr(M_i = 1|Y_{i1}; \boldsymbol{\theta}_t) = 1 - \pi_i$. Then each individual's contribution to the standard GEE is weighted by the inverse probability of being observed, thereby accounting for those subjects with the same history of marker values (and, possibly, other covariates), whose Y_{i2} is missing. Specifically, the GEE in Equation (2.9) is weighted by the inverse probabilities to yield a simple "weighted

GEE"

$$\mathcal{U}_W^{obs}(\boldsymbol{\beta}) = \sum_{i=1}^{N} \left(\frac{I(M_i = 2)}{\pi_i} + \frac{I(M_i = 1)}{1 - \pi_i} \right) \boldsymbol{D}_{i,(m_i)}^{\top} \boldsymbol{V}_{i,(m_i)}^{-1} (\boldsymbol{Y}_{i,(m_i)} - \boldsymbol{\mu}_{i,(m_i)}) = \mathbf{0}. \quad (2.10)$$

By doing so, the observed data are re-weighted to mimic what would likely be observed in a data set without missing data. It can been easily verified that (2.10) is an unbiased estimating equation and, thus, the estimator obtained by solving (2.10) is consistent. One issue, though, is that the dropout probabilities, π_i, are unknown, which means that they should be estimated based on the observed data. Robins et al. (1995) proposed a very general class of "semi-parametrically efficient" weighted estimators for longitudinal data provided that the missingness model has been correctly specified.

A further development, the so-called doubly robust, or doubly protected, estimators (which were introduced in the discussion rejoinder in Scharfstein et al. (1999) and in the contribution to the discussion of Bickel and Kwon (2001) by Robins and Rotnitzky) are robust under certain conditions to misspecification of the model for the probability of dropout. Weighted GEE are consistent when the dropout model is correctly specified, whereas imputation methods are consistent when the imputation model is correctly specified. The key idea behind doubly robust methods is to augment the weighted GEE by a function of the outcome. When this augmentation term is properly selected and modeled correctly according to the distribution of the complete data, the estimator of $\boldsymbol{\beta}$ is consistent even if the model for dropout is misspecified. On the other hand, if the model for missingness is correctly specified, the augmentation term does not need to be correctly specified to yield consistent estimators of $\boldsymbol{\beta}$. Thus, doubly robust estimators are doubly robust in the sense of providing double protection against model misspecification. A more theoretical exposition of doubly robust methods is provided in Tsiatis (2007), whereas a nice introduction to the application of such methods is given by Carpenter et al. (2006).

2.3.3 Random effects models

Another alternative approach for modeling longitudinal marker data is by assuming that the correlation in the data comes from a set of unobserved characteristics attached to each individual, referred to as the random effects. Conditionally on the random effects, $\boldsymbol{b}_i$, which are assigned a prior distribution, e.g. $\boldsymbol{b}_i \sim N(\boldsymbol{0}, \boldsymbol{D})$, the marker measurements are assumed to be independent. Thus, after integrating the random effects out of the marker distribution, a certain model for the correlation between two marker measurements of the same individual is induced, though this may not be available in closed form. The idea of the random effects was initially applied to continuous normally distributed outcomes (Harville, 1977, Laird and Ware, 1982) and the first extensions to more general distributions within the exponential family were provided by West (1985), Wong and Mason (1985), but the concept was popularized by Breslow and Clayton (1993) using the term "generalized linear mixed model" (GLMMs).

More specifically, GLMMs assume that the marker measurements $Y_{ij}|\boldsymbol{b}_i, j = 1, 2, \ldots, Q_i$, are independent given the random effects, with conditional distributions from the exponential family:

$$f(\boldsymbol{Y}_i|\boldsymbol{b}_i; \boldsymbol{\theta}_L) = \prod_{j=1}^{Q_i} \exp\left\{ \frac{Y_{ij}\theta_{ij} - \psi(\theta_{ij})}{\phi} + c(Y_{ij}, \phi) \right\}, \tag{2.11}$$

where ϕ is a dispersion parameter, θ_{ij} is the canonical parameter, and $\psi(\cdot)$ and $c(\cdot)$ are known functions. The conditional mean, $E(Y_{ij}|\boldsymbol{b}_i)$, is related to the canonical parameter through $E(Y_{ij}|\boldsymbol{b}_i) = \psi'(\theta_{ij}) = \frac{\partial\psi(\theta_{ij})}{\partial\theta_{ij}}$, and is further related to the linear predictor through a known link function $g(\cdot)$, i.e. $g\{E(Y_{ij}|\boldsymbol{b}_i)\} = \eta_{ij}$, where $\eta_{ij} = \boldsymbol{x}_{ij}^\top\boldsymbol{\beta} + \boldsymbol{z}_{ij}^\top\boldsymbol{b}_i$. In matrix notation, $\boldsymbol{\eta}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i$, where $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ are the full design matrices of the fixed and random effects, $\boldsymbol{\beta}$ and $\boldsymbol{b}_i$, respectively. Typically, $\boldsymbol{Z}_i$ is a subset of $\boldsymbol{X}_i$ including some function of the measurement times. The conditional variance is equal to $Var(Y_{ij}|\boldsymbol{b}_i) = \phi\psi''(\theta_{ij}) = \phi V\{E(Y_{ij}|\boldsymbol{b}_i)\}$, where $V(\cdot)$ is the variance function.

A critical point that should be clarified is that the regression parameters from the marginal models (e.g. GEE) do not have the same interpretation as the fixed-effects regression parameters in GLMMs since the former models describe the marginal

means whereas the latter the conditional ones (Breslow and Clayton, 1993, Zeger et al., 1988). The key difference between the two modeling frameworks is that the GLMMs include the term $\boldsymbol{z}_{ij}^{\top}\boldsymbol{b}_i$ in the linear predictor, which implies that the fixed effects, $\boldsymbol{\beta}$, do not have a marginal interpretation but rather reflect the effects of covariates conditionally on the random effects. Thus, $\boldsymbol{\beta}$ reflects the effects of covariates when the random effects are equal to zero or comparing two randomly sampled individuals with different covariates but with the same value of the random effects. In principle, an equivalent marginal model derived from a random effects model can be identified through $E(Y_{ij}) = \int g^{-1}(\boldsymbol{x}_{ij}^{\top}\boldsymbol{\beta} + \boldsymbol{z}_{ij}^{\top}\boldsymbol{b}_i)f(\boldsymbol{b}_i)d\boldsymbol{b}_i$, where $f(\boldsymbol{b}_i)$ is the prior distribution for the random effects, typically assumed $\boldsymbol{b}_i \sim N(\boldsymbol{0}, \boldsymbol{D})$. Expect for very special cases, this model is not available in closed form. However, if $g(\cdot)$ is the identity link function, $E(Y_{ij}) = \boldsymbol{x}_{ij}^{\top}\boldsymbol{\beta}$, which implies that the regression coefficients from marginal and random-effects models have the same interpretation. A thorough discussion of the differences between marginal and random-effects models is provided in Zeger et al. (1988).

A standard difficulty with GLMMs is that the estimation process requires integration over the distribution of the random effects, $\int f(\boldsymbol{Y}_i|\boldsymbol{b}_i;\boldsymbol{\theta}_L)f(\boldsymbol{b}_i;\boldsymbol{\theta}_L)d\boldsymbol{b}_i$, which can be time-consuming or can lead to convergence problems. Initial approaches used approximate procedures for inference (Breslow and Clayton, 1993), but nowadays, various improved approaches have been proposed and implemented in user-friendly statistical packages (Capanu et al., 2013, Pinheiro and Bates, 1995).

Due to the independence of Y_{ij} given the random effects, the distribution of the observed marker measurements is easily derived. Thus, given $M_i = m_i$, the observed likelihood of the GLMMs is equal to $\int f(\boldsymbol{Y}_{i,(m_i)}|\boldsymbol{b}_i;\boldsymbol{\theta}_L)f(\boldsymbol{b}_i;\boldsymbol{\theta}_L)d\boldsymbol{b}_i$, where $f(\boldsymbol{Y}_{i,(m_i)}|\boldsymbol{b}_i;\boldsymbol{\theta}_L) = \prod_{j=1}^{m_i}\exp\left\{\frac{Y_{ij}\theta_{ij}-\psi(\theta_{ij})}{\phi} + c(Y_{ij},\phi)\right\}$. In contrast to the standard GEE, GLMMs are valid also under MAR dropout, this requires though that the likelihood model for the marker process is correctly specified.

2.3.3.1 Linear mixed models

Linear mixed models (LMMs) can be considered as a special case of GLMMs using the Normal distribution for the marker, along with the identity link function. A nice motivation for LMMs based on a two-stage procedure is described in Molenberghs and Verbeke (2000). At the first stage, the marker measurements $\boldsymbol{Y}_i$ are assumed to follow the linear regression model:

$$\boldsymbol{Y}_i = \boldsymbol{Z}_i\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i, \tag{2.12}$$

where $\boldsymbol{Z}_i$ is a $Q_i \times q$ matrix of covariates describing how the marker evolves over time for the ith individual. Moreover, $\boldsymbol{\beta}_i$ is a q-dimensional vector of unknown subject-specific regression coefficients, and $\boldsymbol{\epsilon}_i$ is the vector of within-subject residuals assumed to follow independent normal distributions, $\boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \sigma^2\boldsymbol{I}_{Q_i})$. At the second stage, a multivariate regression model of the form

$$\boldsymbol{\beta}_i = \boldsymbol{K}_i\boldsymbol{\beta} + \boldsymbol{b}_i, \tag{2.13}$$

is used to explain the observed variability between subjects with respect to their subject-specific regression coefficients $\boldsymbol{\beta}_i$. $\boldsymbol{K}_i$ is a $(q \times p)$ matrix of known covariates, and $\boldsymbol{\beta}$ is a p-dimensional vector of unknown regression parameters. Finally, the $\boldsymbol{b}_i$ are assumed to be independent, following a q-dimensional Normal distribution with mean vector zero and general positive definite covariance matrix $\boldsymbol{D}$. In practice, the regression parameters in Equation (2.13) are of primary interest. They could be estimated by sequentially fitting the models (2.12) and (2.13). First, all $\boldsymbol{\beta}_i$ are estimated by fitting model (2.12) to the observed data vector $\boldsymbol{Y}_i$ of each subject separately, yielding estimates $\hat{\boldsymbol{\beta}}_i$. Afterwards, model (2.13) is fitted to the estimates $\hat{\boldsymbol{\beta}}_i$, providing inferences for $\boldsymbol{\beta}$.

Obviously, such a two-stage approach would suffer from several problems. First, there may be a severe loss of information when summarizing the marker data $\boldsymbol{Y}_i$ by $\hat{\boldsymbol{\beta}}_i$. This means that random variability is introduced by replacing the $\boldsymbol{\beta}_i$'s in model (2.13) by their estimates $\hat{\boldsymbol{\beta}}_i$, $i = 1, 2, \ldots, N$. Moreover, the covariance matrix of $\hat{\boldsymbol{\beta}}_i$

highly depends on the number of measurements available for the i-th subject, as well as on the time points at which these measurements were taken.

The models can be combined into one unified model by replacing $\hat{\boldsymbol{\beta}}_i$ in Equation 2.12 by the expression in Equation (2.12), yielding

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i,$$

where $\boldsymbol{X}_i = \boldsymbol{Z}_i\boldsymbol{\beta}_i$ is the appropriate $(Q_i \times p)$ matrix of known covariates, and where all other components are as defined earlier. This is an LMM with fixed effects $\boldsymbol{\beta}$ and with subject-specific random effects $\boldsymbol{b}_i$ (Harville, 1977, Laird and Ware, 1982). It is also assumed that $\boldsymbol{b}_i$ and $\boldsymbol{\epsilon}_i$ are independent, with the most common case, $\boldsymbol{b}_i \sim N(\boldsymbol{0}, \boldsymbol{D})$ and $\boldsymbol{\epsilon}_i \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{I}_{Q_i})$. In several approaches in the literature (Bates et al., 2000), a re-parametrization is performed on the covariance matrix of the random effects, $\boldsymbol{b}_i \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{D})$, mainly for computational purposes. It can be easily shown that the marginal distribution of $\boldsymbol{Y}_i$ is multivariate Normal, $\boldsymbol{Y}_i \sim N(\boldsymbol{X}_i\boldsymbol{\beta}, \sigma^2\boldsymbol{V}_i)$, with $\boldsymbol{V}_i = \boldsymbol{I} + \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i^\top$. Thus, it is evident that the fixed effects in an LMM model have a marginal interpretation, coinciding with the interpretation of the regression parameters in the corresponding marginal model (e.g. GEE). One frequently applied special case of LMMs is the random intercept and slope model, which is more suitable when the marker evolves more or less linearly over time. Specifically, it is assumed that

$$Y_{ij} = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1})t_{ij} + \epsilon_{ij},$$

where b_{i0} is frequently referred to as the "random intercept" as it measures the deviation of individual's i baseline marker value from the population baseline (i.e. β_0). The same holds for the random slope, b_{i1}, which measures the difference of the individual's i rate of change relative to the population-averaged rate of change, β_1. Thus, each individual has his/her own intercept and slope. The covariance function of the model is $Cov(Y_{ij}, Y_{ik}) = \sigma^2\{D_{11} + (t_{ij} + t_{ik})D_{21} + t_{ij}t_{ik}D_{22} + 1\}$, which implies that the variance function of the marker is quadratic over time with positive curvature D_{22}. Thus, it is evident that LMMs are fully parametric approaches, and therefore, when there is MAR dropout, the whole marker likelihood model (i.e. both the mean marker evolution and the

random-effects structure) should be correctly specified to ensure consistent parameter estimates. Given the dropout indicators $M_i = m_i$, as $\boldsymbol{Y}_{i,(m_i)} \sim N(\boldsymbol{X}_{i,(m_i)}, \sigma^2 \boldsymbol{V}_{i,(m_i)})$, the observed data likelihood for the marker model assuming MAR dropout is equal to

$$\begin{aligned} \ell(\boldsymbol{\theta}_L) &= \log f(\boldsymbol{Y}_{1,(m_1)}, \ldots, \boldsymbol{Y}_{N,(m_N)}; \boldsymbol{\theta}_L) = -\frac{1}{2} \sum_{i=1}^{N} \log |\sigma^2 \boldsymbol{V}_{i,(m_i)}| \\ &- \frac{1}{2\sigma^2} \sum_{i=1}^{N} (\boldsymbol{Y}_{i,(m_i)} - \boldsymbol{X}_{i,(m_i)} \boldsymbol{\beta})^\top \boldsymbol{V}_{i,(m_i)}^{-1} (\boldsymbol{Y}_{i,(m_i)} - \boldsymbol{X}_{i,(m_i)} \boldsymbol{\beta}). \end{aligned}$$

Then it can be straightforwardly verified that, conditionally on $\boldsymbol{D}$, the estimates for $\boldsymbol{\beta}$ and σ^2, are equal to

$$\begin{aligned} \hat{\boldsymbol{\beta}}(\boldsymbol{D}) &= \left(\sum_{i=1}^{N} \boldsymbol{X}_{i,(m_i)}^\top \boldsymbol{V}_{i,(m_i)}^{-1} \boldsymbol{X}_{i,(m_i)} \right)^{-1} \sum_{i=1}^{N} \boldsymbol{X}_{i,(m_i)}^\top \boldsymbol{V}_{i,(m_i)}^{-1} \boldsymbol{Y}_{i,(m_i)}, \\ \hat{\sigma^2}(\boldsymbol{D}) &= \frac{1}{N} \sum_{i=1}^{N} \left\{ \boldsymbol{Y}_{i,(m_i)} - \boldsymbol{X}_{i,(m_i)} \hat{\boldsymbol{\beta}}(\boldsymbol{D}) \right\}^\top \boldsymbol{V}_{i,(m_i)}^{-1} \left\{ \boldsymbol{Y}_{i,(m_i)} - \boldsymbol{X}_{i,(m_i)} \hat{\boldsymbol{\beta}}(\boldsymbol{D}) \right\}. \end{aligned}$$

Thus, plugging in these solutions into the likelihood function, one obtains a profile-likelihood approach (e.g. Bates et al., 2000), which can be maximized through a numerical optimization method, parameterizing $\boldsymbol{D}$ based on e.g. the Cholesky factorization (Lindstrom and Bates, 1988) or the matrix logarithm scale (Bates et al., 2000, Leonard and Hsu, 1992). The likelihood can be alternatively maximized through the EM algorithm (Dempster et al., 1977), as descibed in McLachlan and Krishnan (1996). Similarly to the general linear model, to account for the loss in degrees of freedom due to estimating $\boldsymbol{\beta}$, one could use the REML estimators, obtained by maximizing $\ell(\boldsymbol{\theta}_L) - \frac{1}{2} \log |\sigma^{-2} \sum_{i=1}^{N} \boldsymbol{X}_{i,(m_i)}^\top \boldsymbol{V}_{i,(m_i)}^{-1} \boldsymbol{X}_{i,(m_i)}|$ (Lindstrom and Bates, 1988).

In many applications, simple random-effects structures (e.g. a random intercept and slope structure) are often employed. However, since every random-effects specification actually implies a certain model for the marginal covariance of the marker, assuming an over-simplistic covariance structure may result in biased population parameter estimates under MAR dropout. Therefore, to improve the estimation of the covariance in the data, two different approaches have been proposed in the literature. In the first one, a high-dimensional vector of functions of time t, expressed in terms of high-order polynomials or splines, can be used in the design matrix of the random effects (Rizopoulos,

2012b). Compared to polynomials, though, splines are usually preferred due to their local natural and better numerical properties (Ruppert et al., 2003); e.g. B-splines have been employed by Rizopoulos et al. (2009)and Brown et al. (2005). An alternative approach is to incorporate an additional stochastic process into the LMM to capture the remaining serial correlation in the observed measurements not captured by the random effects, e.g. the Brownian motion stochastic process (Stirrup et al., 2015, Taylor and Law, 1998). Based on Rizopoulos (2012b), the choice between the two approaches is to a large extent based on the analyst's belief about the true data generating mechanism. More specifically, a random-effects-based only model assumes that the subject-specific marker trajectory is fully determined by the random effects, which is a time-constant inherent characteristic of the individual. On the other hand, models including both random effects and additional stochastic processes attempt to more precisely capture the features of the marker trajectory by allowing the subject-specific trends to vary in time. It should be emphasized that as both the random-effects and the stochastic serial correlation approaches attempt to appropriately model the marginal correlation in the data, there is a contest for information between the two procedures. Thus, a model including an elaborate random-effects structure and a model with a simpler one but with an additional stochastic process may result in indistinguishable fits to the data (Rizopoulos, 2012b).

We now briefly describe the basics of Brownian motion, a stochastic process that is often used in LMMs to model the remaining serial correlation in the observed measurements not captured by the random effects. Brownian motion is a non-stationary stochastic process constituting a continuous-time generalization of a simple random walk (Grimmett et al., 2001), in which successive increments are independent of the history of the process. When considered in terms of a given set of measurement times, though, a scaled Brownian motion process, denoted by W_t at time t, is defined by the following equations:

$$\begin{aligned} W_0 &= 0 \\ W_t - W_s &\sim N\{0, \kappa(t-s)\} \quad \text{for} \quad 0 \leq s < t, 0 < \kappa. \end{aligned}$$

The baseline ($t = 0$) value of the process is equal to 0 with its increments being stationary, independent (for disjoint periods of time) and normally distributed with mean zero and variance equal to the difference in time between observation points scaled by a positive constant factor κ. Therefore, the following properties are implied

$$\begin{aligned} E(W_t) &= 0, \\ Var(W_t) &= \kappa t, \\ Cov(W_s, W_t) &= \kappa \min(s, t). \end{aligned}$$

Thus, the distribution of a set of n observations at given time points follows a multivariate Normal distribution with zero mean and covariance matrix as defined by the previous formulas. For example, a scaled Brownian motion can be incorporated in a random intercept and slope model as follows:

$$Y_i(t) = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1})t + W_t + \epsilon_i(t).$$

Since the baseline ($t = 0$) value of the scaled Brownian motion process is 0, the random intercept and the within-individual variance are the only sources of variability at baseline. For times different from zero, the covariance structure of the marker process becomes more complex as it depends on the covariance function of depends on the $\{W_t\}_{t=0}^{\infty}$.

In some cases a scaled Brownian motion may not suffice to appropriate describe the correlation in the data not captured by the random effects. In such cases, the Fractional Brownian motion process can be used, which represents an extension of the scaled Brownian motion process. In the fractional Brownian motion process, the increments for disjoint time periods are not constrained to be independent, though they do remain stationary. The process was firstly introduced by Mandelbrot and Ness (1968). The characteristics of a fractional Brownian motion process are determined by one additional parameter, referred to as H or "the Hurst index", that taking a value in (0,1). Standard Brownian motion constitutes a special case of fractional Brownian motion, corresponding to $H = 1/2$. As for standard Brownian motion, the expectation

of the value of the process is zero for all points in time. When $H < 1/2$, successive increments of the process are negatively correlated, whereas, for $H > 1/2$, successive increments of the process are positively correlated. As for Brownian motion, a positive scale parameter (κ) can be added to the standard definition of fractional Brownian motion, corresponding to the variance of the process at $t = 1$. We may then characterize the properties of the process as follows:

$$\begin{aligned} W_0 &= 0, \\ E(W_t) &= 0, \\ Var(W_t) &= \kappa|t|^{2H}, \\ Cov(W_s, W_t) &= \frac{\kappa}{2}(|s|^{2H} + |t|^{2H} - |t-s|^{2H}). \end{aligned}$$

Similarly, fractional Brownian motion, when evaluated at a given set of time points, a multivariate Normal distribution is implied. Thus, (fractional and non-fractional) Brownian motion processes seem to have very appealing properties, and as shown by Stirrup et al. (2015), they can be readily incorporated into the theoretical framework of LMMs.

2.4 Joint modeling of longitudinal and time-to-dropout data

In this section, we describe the most common approaches for joint modeling of marker and missing data process, i.e. the selection, pattern-mixture, and shared-parameter models.

Selection models

In selection models, one first assumes a complete data model for the marker, which characterizes the evolution of the marker that would have been observed in the absence of dropout. Then a model for the probability of dropout at each time point is assumed to depend not only on the observed marker values up to that point but also on future

values that would have obtained after dropout. That is, selection models specify the joint distribution of M_i and $\boldsymbol{Y}_i$ via models for the marginal distribution of $\boldsymbol{Y}_i$ and the conditional distribution of M_i given $\boldsymbol{Y}_i$, i.e.

$$f(\boldsymbol{Y}_i, M_i; \boldsymbol{\theta}_L, \boldsymbol{\theta}_t) = f(\boldsymbol{Y}_i; \boldsymbol{\theta}_L) \Pr(M_i|\boldsymbol{Y}_i; \boldsymbol{\theta}_t), \tag{2.14}$$

with the dropout mechanism $\Pr(M_i|\boldsymbol{Y}_i; \boldsymbol{\theta}_t)$ specifying dependence on the missing marker values on top of the observed ones. For identifiability, though, the models are usually restricted in some way by specifying parametric assumptions about $f(\boldsymbol{Y}_i; \boldsymbol{\theta}_L)$ and models for the dependence of the dropout probabilities on the unobserved outcomes.

Early development of selection models, especially in the econometric literature, was due to the Tobit model (Heckman, 1976). This approach combines a marginal Normal regression model for the marker with a Normal-based random threshold model for the missingness probability. For example, $Y \sim N(\mu, \sigma^2)$, with the probability of a value being missing depending on a different Normal random variable $Y_m \sim N(\mu_m, \sigma_m^2)$, where $\Pr(R = 0) = \Pr(Y_m < 0)$. To introduce dependence of missingness on Y, Y and Y_m are assumed to be correlated. To avoid the potentially complex direct likelihood maximization, Heckman (1976) used a two-step inferential procedure. It should be noted that the use of the Tobit model led to considerable debate in the econometric literature, mainly raising issues about the identifiability and sensitivity of the model (Amemiya, 1984). At first, the Tobit model does not appear to lie within the selection-models family specified in Equation (2.14). However, due to the bivariate Normal distribution of (Y, Y_m), it is straightforward to show that in the Tobit model, $\Pr(R = 0|Y = y) = \Phi(\beta_0 + \beta_1 y)$, with suitably chosen parameters β_0 and β_1 and $\Phi()$ denoting the Normal cumulative distribution function. This can be seen as a probit regression model for the (binary) missingness indicators.

This basic structure has formed the basis for more advanced selection models proposed for longitudinal data under dropout. For example, Diggle and Kenward (1994) proposed a selection model combining a multivariate Normal model for the complete marker data, $\boldsymbol{Y}_i$, with a logistic regression model for the dropout probabilities conditionally on both observed and unobserved marker data. For example, based on Molen-

berghs and Verbeke (2000), a simplified version of the model proposed by Diggle and Kenward (1994) is

$$\Pr(M_i = j | M_i \geq j, \boldsymbol{Y}_i; \boldsymbol{\theta}_t) = \theta_{t0} + \theta_{t1} Y_{ij} + \theta_{t2} Y_{ij+1}, \tag{2.15}$$

where $\Pr(M_i = j | M_i \geq j, \boldsymbol{Y}_i; \boldsymbol{\theta}_t)$ denotes the hazard for dropout (with $Y_{ij+1}, \ldots, Y_{iQ_i}$ being missing given $M_i = j$). Thus, under this model, dependence of dropout on the observed and unobserved marker values is measured through the parameters, θ_{t1} and θ_{t2}, respectively. Note that $\theta_{t1} = \theta_{t2} = 0$ implies MCAR dropout and if $\theta_{t2} \neq 0$, dropout is MNAR, whereas MAR is a special case where $\theta_{t2} = 0$ and $\theta_{t1} \neq 0$. Thus, under the assumed model, it is possible to test for nonignorable dropout by testing the hypothesis H0: $\theta_{t2} = 0$, e.g. through a likelihood ratio test, though it has been shown that such tests are sensitive to minor violations of the model assumptions (Molenberghs and Verbeke, 2000).

The model in Equation (2.15) relates dropout to the current and previous observation only, though Diggle and Kenward (1994) considered a more general model where dropout at occasion $j + 1$ (i.e. $M_i = j$) can depend on the complete observed history $\{Y_{i1}, \ldots, Y_{ij}\}$, as well as on external covariates. In theory, selection models could allow dependence of dropout on all future observed marker values. However, including all future marker measurements in Equation (2.15) (i) may be counter-intuitive in many cases, especially when it is difficult for the study participants to make projections about the future responses (Molenberghs and Verbeke, 2000) and (ii) leads to an observed data likelihood that requires evaluating cumbersome high-dimensional integrals over the missing values. Diggle and Kenward (1994) used the simplex algorithm (Nelder and Mead, 1965) to maximize the likelihood, and Molenberghs et al. (1997) proposed a similar selection model carried out through the EM algorithm (Dempster et al., 1977).

The main advantage of selection models for informative dropout is that they directly model the quantities of primary interest, i.e. the marginal distribution of the marker process $\boldsymbol{Y}_i$ and the distribution of the dropout process conditional on $\boldsymbol{Y}_i$. The former is used for marginal inferences on marker profiles; the latter is used to characterize the

nature of the dropout process (MCAR, MAR, MNAR). Several authors have pointed out that results from such models may be highly sensitive to the model's assumptions. For example, Kenward (1998) reanalyzed the data of Diggle and Kenward (1994) and found that all evidence for MNAR dropout vanishes when the normality assumption is replaced by a heavy-tailed t-distribution. On top of that, the model of Diggle and Kenward (1994) is easier to comprehend and conceptualize when marker data are collected at fixed equally-spaced visits and it is plausible to consider marker values after the dropout event. In human studies, though, such a strict visiting protocol is hardly the case and the exact time of dropout (e.g. time to death or disease progression) may be of particular interest. These difficulties with the selection model proposed by Diggle and Kenward (1994) have motivated the use of alternative approaches in which the marker and dropout processes are linked through common parameters, e.g. the marker subject-specific random effects, hence the term "shared random effects models" (SREMs), described and discussed below:

Shared random effects models

In SREMs, it is assumed that both $\boldsymbol{Y}_i|\boldsymbol{b}_i$ and $M_i|\boldsymbol{b}_i$ depend on shared latent variables, $\boldsymbol{b}_i$. More specifically, the model for $\boldsymbol{Y}_i$ is linked with a model for M_i via a vector of random effects that are shared between the complete data model and the model for the dropout mechanism. Conceptually, each individual is assumed to have subject-specific random effects influencing both each marker value, Y_{ij}, and the probability of dropout, M_i. Note also that, for such models, Wu and Bailey (1989) used the term right informative censoring models. However, Little (1995) preferred the term non-ignorable random-coefficient-based dropout to distinguish this type of models from those of the form of Diggle and Kenward (1994), called non-ignorable outcome-based dropout models, as in the latter, it is assumed that dropout depends on the hypothetical marker measurements collected after dropout. The complete data likelihood of these models is equal to

$$f(\boldsymbol{Y}_i, M_i, \boldsymbol{b}_i; \boldsymbol{\theta}) = f(\boldsymbol{Y}_i|\boldsymbol{b}_i; \boldsymbol{\theta}) f(\boldsymbol{b}_i; \boldsymbol{\theta}) \Pr(M_i|\boldsymbol{b}_i, \boldsymbol{Y}_i; \boldsymbol{\theta}),$$

where $\boldsymbol{\theta}^\top = (\boldsymbol{\theta}_L^\top, \boldsymbol{\theta}_t^\top)$ denotes the full parameter vector of the model. In most applications, though, $\boldsymbol{Y}_i|\boldsymbol{b}_i$ and $M_i|\boldsymbol{b}_i$ are assumed to be independent of one another, thus,

$$f(\boldsymbol{Y}_i, M_i, \boldsymbol{b}_i; \boldsymbol{\theta}) = f(\boldsymbol{Y}_i|\boldsymbol{b}_i; \boldsymbol{\theta}) f(\boldsymbol{b}_i; \boldsymbol{\theta}) \Pr(M_i|\boldsymbol{b}_i; \boldsymbol{\theta}),$$

though extended SREMs relaxing the conditional independence assumption have also been proposed in the literature (Creemers et al., 2011, Njagi et al., 2014), mainly constituting tools for sensitivity analyses adding unidentifiable random terms. Recall that given $M_i = m_i$, $\boldsymbol{Y}_i^\top = (\boldsymbol{Y}_{i,(m_i)}^\top, \boldsymbol{Y}_{i,(\overline{\boldsymbol{m}_i})}^\top)$. It is, thus, evident that the observed data likelihood of SREMs requires integration over both the latent variables, $\boldsymbol{b}_i$, and the missing marker values, $\boldsymbol{Y}_{i,(\overline{\boldsymbol{m}_i})}$, i.e.

$$\begin{aligned} f(\boldsymbol{Y}_{i,(m_i)}, M_i; \boldsymbol{\theta}) &= \int\int f(\boldsymbol{Y}_i|\boldsymbol{b}_i; \boldsymbol{\theta}) f(\boldsymbol{b}_i; \boldsymbol{\theta}) \Pr(M_i|\boldsymbol{b}_i; \boldsymbol{\theta}) d\boldsymbol{Y}_{i,(\overline{\boldsymbol{m}_i})} d\boldsymbol{b}_i \\ &= \int\int f(\boldsymbol{Y}_i|\boldsymbol{b}_i; \boldsymbol{\theta}) d\boldsymbol{Y}_{i,(\overline{\boldsymbol{m}_i})} f(\boldsymbol{b}_i; \boldsymbol{\theta}) \Pr(M_i|\boldsymbol{b}_i; \boldsymbol{\theta}) d\boldsymbol{b}_i \\ &= \int f(\boldsymbol{Y}_{i,(m_i)}|\boldsymbol{b}_i; \boldsymbol{\theta}) f(\boldsymbol{b}_i; \boldsymbol{\theta}) \Pr(M_i|\boldsymbol{b}_i; \boldsymbol{\theta}) d\boldsymbol{b}_i, \end{aligned} \tag{2.16}$$

where $f(\boldsymbol{Y}_{i,(m_i)}|\boldsymbol{b}_i; \boldsymbol{\theta})$ is the marginal distribution of the observed marker values conditionally on the random effects. As in most applications, $Y_{ij}|\boldsymbol{b}_i$, $j = 1, 2, \ldots, Q_i$ are assumed to be independent given the random effects, the distribution of $f(\boldsymbol{Y}_{i,(m_i)}|\boldsymbol{b}_i; \boldsymbol{\theta})$ is easy to obtain. As also pointed out by Rizopoulos (2012b), the observed data likelihood of the SREMs is easily obtained without requiring explicit integration with respect to the missing marker values even under intermittent missingness. This is in contrast to the selection and pattern-mixture model framework where cumbersome computations are needed to evaluate the likelihood (Jansen and Molenberghs, 2008, Troxel et al., 1998a,b).

Because the vector of the random effects, $\boldsymbol{b}_i$, is shared between the two models, marginally, i.e. after integrating over the distribution of the random effects, this induces correlation between M_i and the observed and unobserved components of $\boldsymbol{Y}_i$. Thus, the resulting missing data mechanism is MNAR, though the underlying mechanism for the dependence of M_i on $\boldsymbol{Y}_i$ may not be immediately transparent. To clarify this, as

discussed in Rizopoulos (2012b), note the dropout mechanism implied by SREMs is equal to

$$\Pr(M_i = m_i|\boldsymbol{Y}_{i,(m_i)}, \boldsymbol{Y}_{i,(\overline{\boldsymbol{m}_i})}; \boldsymbol{\theta}) = \int \Pr(M_i = m_i|\boldsymbol{b}_i, \boldsymbol{Y}_{i,(m_i)}, \boldsymbol{Y}_{i,(\overline{\boldsymbol{m}_i})}; \boldsymbol{\theta})$$
$$\times f(\boldsymbol{b}_i|\boldsymbol{Y}_{i,(m_i)}, \boldsymbol{Y}_{i,(\overline{\boldsymbol{m}_i})}; \boldsymbol{\theta}) d\boldsymbol{b}_i = \int \Pr(M_i = m_i|\boldsymbol{b}_i; \boldsymbol{\theta}) f(\boldsymbol{b}_i|\boldsymbol{Y}_{i,(m_i)}, \boldsymbol{Y}_{i,(\overline{\boldsymbol{m}_i})}; \boldsymbol{\theta}) d\boldsymbol{b}_i, \quad (2.17)$$

where the simplification in the last line is due to the conditional independence assumption. Thus, it is evident that the dependence of dropout on the missing observation is induced through the posterior distribution of the random effects, $f(\boldsymbol{b}_i|\boldsymbol{Y}_{i,(m_i)}, \boldsymbol{Y}_{i,(\overline{\boldsymbol{m}_i})}; \boldsymbol{\theta})$. It should be emphasized, though, that if there is no dependence of M_i on the random effects, $\boldsymbol{b}_i$, then $\Pr(M_i = m_i|\boldsymbol{b}_i; \boldsymbol{\theta})$ can be moved outside the integral sign, and thus $\Pr(M_i = m_i|\boldsymbol{Y}_{i,(m_i)}, \boldsymbol{Y}_{i,(\overline{\boldsymbol{m}_i})}; \boldsymbol{\theta}) = \Pr(M_i = m_i; \boldsymbol{\theta})$ as $\int f(\boldsymbol{b}_i|\boldsymbol{Y}_{i,(m_i)}, \boldsymbol{Y}_{i,(\overline{\boldsymbol{m}_i})}; \boldsymbol{\theta}) d\boldsymbol{b}_i = 1$, which implies MCAR dropout (or covariate-dependent missingness if some covariates are considered in the dropout model). Therefore, in the SREMs, the dropout mechanism is allowed to be either MNAR or MCAR; that is, MAR cannot hold without reducing to MCAR (Njagi et al., 2014). Although SREMs are often treated as a separate category in the literature, as Equation (2.17) revealed, they can be regarded as special cases of the selection models, but they can also, theoretically, be represented within the pattern-mixture models family (Molenberghs et al., 2014). Some frequently used specific SREM models are outlined below.

Much of the early development in the field was due to Wu and Carroll (1988) who modeled the disease marker by an LMM and a probit model for the dropout process depending on both the "true" (i.e. including both fixed and random effects) initial value and slope of the marker and on discrete time components. They also derived a likelihood ratio test to test for informative dropout. Moreover, Wu and Bailey (1988, 1989) showed that under this probit model, the expected individual slope is a monotonic function of the dropout time. However, a limitation of these approaches is that they assumed that all patients are potentially followed up the same length of time, i.e., staggered entry was not allowed. On top of that, patients with only a single marker measurement for whom a slope cannot be calculated must be excluded from the analysis.

In most real data applications in biomedical studies, though, the time to dropout is continuous. Thus, it is more natural to model the dropout time, say $T_i^\star$, through common continuous-time survival models. Schluchter (1992) extended and formalized the conditional linear model of Wu and Carroll (1988) by modeling jointly the marker measurements and the individuals' survival times and obtained inferences through direct maximum likelihood estimation, which enabled them to take into account all data and to allow for staggered entry. It was assumed that the underlying individual intercept and slope and the log-survival time follow a trivariate Normal distribution. A similar approach was applied by Gruttola and Tu (1994), which is essentially a re-parameterization of the model by Schluchter (1992), but allowing for a more general class of models. Specifically, the model of Gruttola and Tu (1994), termed LN-SREM(RE) from now on, has the following form

$$\begin{aligned} \boldsymbol{Y}_i &= \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i \\ U_i &= \boldsymbol{\gamma}^\top \boldsymbol{w}_i + \boldsymbol{\alpha}^\top \boldsymbol{b}_i + r_i, \end{aligned} \tag{2.18}$$

i.e. the marker process is described through a general LMM, whereas, regarding the log-time to dropout, U_i, $\boldsymbol{w}_i$ denotes baseline covariates with an associated parameter vector, $\boldsymbol{\gamma}$, $\boldsymbol{\alpha}$ denotes a parameter vector that measures the effects of each random effect on the dropout time, and $r_i \sim N(0, s^2)$ the corresponding residuals. Both the approaches proposed by Schluchter (1992) and Gruttola and Tu (1994) applied the EM algorithm to maximize the likelihood. However, both approaches may suffer from convergence issues. Touloumi et al. (1999) proposed a similar model to jointly model a disease marker through a general LMM and the log-time to dropout by assuming that the residuals of the dropout model and the marker random effects follow the multivariate Normal distribution. Estimation was obtained through iterative use of the restricted iterative generalized least squares (RIGLS) estimator along with a nested EM algorithm to incorporate survival censored data, which was computationally less complex than direct maximization of the joint likelihood.

In a considerable amount of the literature on SREMs, the main interest lies in the

time-to-dropout process (e.g. death, drop out of the study, or treatment discontinuation), which terminates the collection of marker measurements, aiming to describe the effects of the "true" marker process on the probability of dropout. The motivation for this comes from the fact that simply including the observed marker values as a time-varying covariate in a survival model can severely underestimate the true association (Prentice, 1982, Wulfsohn and Tsiatis, 1997). Furthermore, two-stage approaches (e.g. Tsiatis et al., 1995) can lead to biased results and are not fully efficient (Wulfsohn and Tsiatis, 1997). For normally distributed markers, perhaps the most frequently applied approach is to assume an LMM for the marker process and a proportional hazards model for the time to dropout conditionally on the "true" (i.e. after eliminating the measurement error) marker value (e.g. Faucett and Thomas, 1996, Mauff et al., 2020, Rizopoulos, 2012a,b, Tsiatis et al., 1995, Wulfsohn and Tsiatis, 1997). Under this setting, it is convenient to reformulate the LMM as $\boldsymbol{Y}_i(t) = \boldsymbol{X}_i^\top(t)\boldsymbol{\beta} + \boldsymbol{Z}_i^\top(t)\boldsymbol{b}_i + \boldsymbol{\epsilon}_i(t)$, where $\boldsymbol{Y}_i(t)$, $\boldsymbol{X}_i(t)$, $\boldsymbol{Z}_i(t)$, and $\boldsymbol{\epsilon}_i(t) \sim N(0, \sigma^2)$ are the observed marker value, the design matrix for the fixed effects, the design matrix of the random effects, and the measurement error at time t since time origin, respectively. Then $m_i(t) = \boldsymbol{X}_i^\top(t)\boldsymbol{\beta} + \boldsymbol{Z}_i^\top(t)\boldsymbol{b}_i$ is interpreted as the "true" marker value, i.e.. the marker value which would have been observed had the measurement error been eliminated.

To quantify the strength of the association between the "true" marker value and the risk for the dropout event, the most commonly applied model is a relative risk model of the form:

$$\begin{aligned} h_i\{t|M_i(t);\boldsymbol{\theta}_t\} &= \lim_{dt\to 0}\frac{\Pr\{t \le T_i^\star < t+dt | T_i^\star \ge t, M_i(t);\boldsymbol{\theta}_t\}}{dt} \\ &= h_0(t;\boldsymbol{\psi})\exp\{\boldsymbol{\gamma}^\top\boldsymbol{w}_i + \alpha m_i(t)\}, \end{aligned} \tag{2.19}$$

where $M_i(t) = \{m_i(s), 0 \le s < t\}$ denotes the history of the "true" unobserved marker process up to time point t, $h_0(t;\boldsymbol{\psi})$ denotes the baseline hazard function, and $\boldsymbol{w}_i$ is a vector of baseline covariates with a corresponding parameter vector, $\boldsymbol{\gamma}$. Similarly, parameter α quantifies the effect of the underlying marker value to the risk for an event, with $\exp(\alpha)$ denoting the relative increase in the hazard for dropout at time t resulting

from one unit increase in $m_i(t)$ at the same time point. Since the model in Equation (2.19) is a proportional hazards model using the current-value parameterization, we refer to it as the PH-SREM(CV) model. Moreover, note that the relative risk model in Equation (2.19) postulates that the risk for an event at time t depends only on the current value of the time-dependent marker $m_i(t)$. However, this does not hold for the survival function as

$$\begin{aligned} S_i\{t|M_i(t);\boldsymbol{\theta}_t\} &= \Pr\{T_i^\star > t|M_i(t);\boldsymbol{\theta}_t\} \\ &= \exp\left[-\int_0^t h_0\{s;\boldsymbol{\psi}\}\exp\{\boldsymbol{\gamma}^\top\boldsymbol{w}_i + \alpha m_i(s)\}ds\right], \end{aligned} \tag{2.20}$$

which implies that the corresponding survival function, in contrast to Equation (2.19), depends on the whole true marker history, $M_i(t)$.

The literature on joint modeling has greatly expanded, including various extensions to the standard parameterization in Equation (2.19). A comprehensive discussion of the available models is provided in Rizopoulos (2012b). One extension is motivated by the fact that the standard parameterization in Equation (2.19) assumes a homogeneous effect of the true marker value on the whole population. However, this is evidently a strong assumption that may not be met in practice. A straightforward extension to handle such situations is to include in the linear predictor of the relative risk model, interaction terms of the marker with baseline covariates of interest (Rizopoulos, 2012b), i.e.,

$$h_i\{t|M_i(t);\boldsymbol{\theta}_t\} = h_0(t;\boldsymbol{\psi})\exp\left[\boldsymbol{\gamma}^\top\boldsymbol{w}_{1i} + \boldsymbol{\alpha}^\top\{\boldsymbol{w}_{2i}m_i(t)\}\right], \tag{2.21}$$

where $\boldsymbol{w}_{1i}$ is used to accommodate the direct effects of baseline covariate to the risk for an event, and $\boldsymbol{w}_{2i}$ contains interaction terms that expand the association of $m_i(t)$ in different subgroups in the data.

In some occasions, the standard assumption that the current value of the "true" marker value affects the current risk for dropout may lead to medically illogical conclusions. A typical example has been presented by Cavender et al. (1992) who, in a study on patients with coronary artery disease, noted that current smoking decreased

the risk for death (though not statistically significantly). This unexpected result may be attributed to the fact that most patients who died were smokers but many of them had stopped smoking at the last visit prior to death. Thus, many of the patients who died had just quit smoking, whereas some of the patients who were still alive were still smoking, leading to the surprising result. To adjust for that, one could apply the time-lagged parameterization. Specifically, one could use the following re-formulation of the model

$$h_i\{t|M_i(t);\boldsymbol{\theta}_t\} = h_0(t;\boldsymbol{\psi})\exp\left[\boldsymbol{\gamma}^\top \boldsymbol{w}_i + \alpha m_i\{max(t-c,0)\}\right], \tag{2.22}$$

which postulates that the hazard for dropout at time t depends on the true value of the marker at time $t-c$, where c denotes the time lag of interest, typically specified by the analyst on medical grounds.

In the previous parameterizations we have assumed that the risk for an event depends on either the current or a previous value of the marker. However, there may be cases where, especially when the marker follows a non-linear evolution over time, it is reasonable to allow the risk for dropout to depend on other features of the marker trajectory. Such a parameterization was considered by Ye et al. (2008) who postulated a joint model in which the risk depends on both the current true value of the marker and the corresponding slope at time t:

$$h_i\{t|M_i(t);\boldsymbol{\theta}_t\} \;=\; h_0(t;\boldsymbol{\psi})\exp\{\boldsymbol{\gamma}^\top \boldsymbol{w}_i + \alpha_1 m_i(t) + \alpha_2 m_i'(t)\}, \tag{2.23}$$

where $m_i'(t) = \frac{\partial \boldsymbol{X}_i^\top(t)}{\partial t}\boldsymbol{\beta} + \frac{\partial \boldsymbol{Z}_i^\top(t)}{\partial t}\boldsymbol{b}_i$. The interpretation of parameter α_1 remains the same as in the standard parameterization (2.19), but comparing individuals with the same marker slope. Parameter α_2 measures the strength of the association between the hazard for dropout and the marker slope, providing that $m_i(t)$ remains constant. As discussed in Rizopoulos (2012b), such parameterizations capture situations where, for example, at a specific time point two patients show similar true marker levels, but they may differ in the rate of change of the marker. Some authors, though, have pointed out that the results based on the time-dependent slope parameterization in Equation (2.23) may be more sensitive to violations of the model assumptions than the results based on

the more standard, current-value parameterization in Equation (2.19) (Crowther et al., 2016, García-Hernandez et al., 2020).

Another type of parameterization that is frequently used in joint models includes only the random effects in the linear predictor of the dropout model, i.e.,

$$h_i\{t|M_i(t);\boldsymbol{\theta}_t\} \quad = \quad h_0(t;\boldsymbol{\psi})\exp(\boldsymbol{\gamma}^\top \boldsymbol{w}_i + \boldsymbol{\alpha}^\top \boldsymbol{b}_i), \tag{2.24}$$

with this model termed PH-SREM(RE) from now on; $\boldsymbol{\alpha}$ denotes a vector of association parameters with each one measuring the association between the corresponding random effect and the hazard for dropout. However, this parameterization is more meaningful when a simple random intercepts and slopes structure is assumed for the marker, in which case, the random effects express subject-specific deviations from the average intercept and average slope, $h_i\{t|M_i(t);\boldsymbol{\theta}_t\} = h_0(t;\boldsymbol{\psi})\exp(\alpha_0 b_{i0} + \alpha_1 b_{i1})$. Then if $\alpha_0 < 0$ and $\alpha_1 < 0$, patients with lower marker values (i.e. $b_{i0} < 0$) at baseline and slopes lower than the population average ($b_{i1} < 0$) have higher probability of dropout. The PH-SREM(RE) model shares some similarities with the LN-SREM(RE) model in the sense that they directly measure the effects of the random effects on the time to dropout. Also, PH-SREM(RE) can be similar to the time-dependent slope parameterization (2.23) under certain models for the marker evolution.

So far, the most commonly applied estimating methods for joint models is the (semi-parametric) maximum likelihood (e.g. Henderson et al., 2000, Hsieh et al., 2006, Wulfsohn and Tsiatis, 1997), with a theoretical treatment of the asymptotic properties of the joint model estimators provided by Zeng and Cai (2005). Nevertheless, Bayesian inference on joint models has been investigated by many authors (Chi and Ibrahim, 2006, Hanson et al., 2011, R. Brown and G. Ibrahim, 2003, Wang and Taylor, 2001), as well.

One standard disadvantage of joint modeling is that the estimation process is quite complex, leading to high computational time under both Frequentist approaches (e.g. maximum likelihood) and Bayesian approaches (e.g. MCMC), when the sample size is large. However, due to recent advances in computational algorithms (Crowther et al., 2012, Rizopoulos, 2010, 2012a, 2016) and the availability of user-friendly statistical

software (Crowther, 2012, Rizopoulos, 2010, 2016), the application of joint models to real data has grown considerably.

We now introduce the likelihood function of SREM models. To take potential right censoring into account, let $T_i = \min(T_i^\star, C_i)$ denote the observed survival time, i.e. the minimum of the true survival time $T_i^\star$ and a hypothetical right censoring time C_i, and $\delta_i = I(T_i^\star < C_i)$ denote the corresponding dropout indicator. No marker measurements are observed after T_i, and it is assumed that there are $M_i = m_i$ observed measurements in total. Thus, the likelihood function for the observed data is equal to

$$f(\boldsymbol{Y}_{i,(m_i)}, T_i, \delta_i; \boldsymbol{\theta}) = \int f(\boldsymbol{Y}_{i,(m_i)}|\boldsymbol{b}_i; \boldsymbol{\theta}) f(\boldsymbol{b}_i; \boldsymbol{\theta}) f(T_i, \delta_i|\boldsymbol{b}_i; \boldsymbol{\theta}) d\boldsymbol{b}_i, \tag{2.25}$$

with the difference between (2.16) and (2.25) being that $\Pr(M_i|\boldsymbol{b}_i; \boldsymbol{\theta})$ has been replaced by $f(T_i, \delta_i|\boldsymbol{b}_i; \boldsymbol{\theta})$, as dropout is modeled in continuous time. Assuming that censoring is non-informative (Kalbfleisch and Prentice, 2002), in the sense that, given the observed history of information, censoring is independent of the true dropout time, the future marker measurements, and the random effects, it can be shown that, for the PH-SREM(CV) model, the likelihood for the ith subject is equal to

$$\begin{aligned} f(T_i, \delta_i|\boldsymbol{b}_i; \boldsymbol{\theta}) &= h_i\{T_i|M_i(T_i); \boldsymbol{\theta}_t\}^{\delta_i} S_i\{T_i|M_i(T_i); \boldsymbol{\theta}_t\} \\ &= h_0\{T_i; \boldsymbol{\psi}\} \exp\{\boldsymbol{\gamma}^\top \boldsymbol{w}_i + \alpha m_i(T_i)\}^{\delta_i} \\ &\times \exp\left[-\int_0^{T_i} h_0\{s; \boldsymbol{\psi}\} \exp\{\boldsymbol{\gamma}^\top \boldsymbol{w}_i + \alpha m_i(s)\} ds\right], \end{aligned}$$

i.e. the standard form of a survival likelihood under non-informative right censoring. Since the integral in Equation (2.25) does not have a closed-form solution, except for very special cases, computationally-intensive methods are, in general, required to approximate the integral. For this reason, early approaches relied on two-step inferential procedures (e.g. Self and Pawitan, 1992, Tsiatis et al., 1995). However, it has been shown that such approaches can lead to biased results (Dafni and Tsiatis, 1998, Tsiatis and Davidian, 2001, Ye et al., 2008). For these reasons, most authors have now focused on full likelihood approaches.

To employ a full-likelihood approach, though, one needs to make assumptions about the baseline hazard, $h_0(t; \boldsymbol{\psi})$. The most convenient method would be to leave it unspec-

ified and use the Cox partial likelihood approach to estimate the parameters. However, as explained in Hsieh et al. (2006), this is not possible within the SREM framework, where a full likelihood approach is required. When a semi-parametric approach is employed, the cumulative hazard function, $H_i(t) = \int_0^t h(s)ds$, is replaced by a step function with jumps at the observed dropout times, leading to a high-dimensional parameter vector causing complications when inverting the Hessian matrix to derive standard errors. Similarly to Cox partial likelihood, a profile likelihood approach could be used, but Hsieh et al. (2006) showed that standard asymptotic results do not apply and proposed Bootstrapping to derive standard errors. Alternative semi-parametric approaches are described in Kim et al. (2017), R. Brown and G. Ibrahim (2003), Song and Wang (2008). Many authors, though, have modeled $h_i(t)$ through flexible parametric models, e.g. B-splines (Rizopoulos, 2012b), restricted cubic splines (Crowther et al., 2012), and piecewise constant functions (Rizopoulos, 2012a). The main computational burden to evaluate the likelihood of the SREM joint models is due to the two integrals involved in Equations (2.25) and (2.20). The former, though, is computationally more demanding as it possibly involves multi-dimensional integration over the distribution of the random effects, whereas, the latter, as it is uni-dimensional, can be efficiently approximated by e.g. the 7-point or 15-point Gauss-Kronrod rule (Press et al., 2007). To evaluate the integral over the random effects, (adaptive or non-adaptive) Gaussian quadrature rules have been typically applied (Crowther et al., 2012, 2016, Henderson et al., 2000, Wulfsohn and Tsiatis, 1997).

The literature in SREMs has substantially grown in the last years, including many alternative parameterizations and estimating approaches. One of the most interesting extensions is to allow for competing risks, i.e. multiple mutually exclusive events causing termination of marker measurements. For example, HIV-positive individuals receiving ART can die while in care or disengage from care, which are competing risks causing termination of collection of CD4 cell count data. Various competing risks SREMs have been proposed recently (e.g. Andrinopoulou et al., 2014, 2017, Elashoff et al., 2008, Hickey et al., 2018, Hu et al., 2009, Huang et al., 2010, Proust-Lima et al.,

2016, Williamson et al., 2008). In principle, competing risks data can be analyzed through either cause-specific hazards or cumulative incidence functions (CIFs). The former measures the rate of failure from a particular cause at a specific time point given that the individual has survived up to that point, whereas the latter is the probability of occurrence of a specific cause over time (Bakoyannis and Touloumi, 2012, Beyersmann et al., 2011, Kalbfleisch and Prentice, 2002). When one wishes to address aetiological-type research questions, cause-specific hazard models seem to be more appropriate, whereas a CIF-based statistical analysis should be used for evaluating the prognosis of a disease, interventions in populations, as well as for prediction purposes in general. In most cases, though, the competing risks submodels are specified in terms of the cause-specific hazards under the SREM framework (Andrinopoulou et al., 2014, 2017, Elashoff et al., 2008, Hu et al., 2009, Rizopoulos, 2012b, Williamson et al., 2008). For example, the survival submodels in terms of the cause-specific hazards using the current-value parameterization are equal to

$$\begin{aligned}\alpha_{ik}\{t|m_i(t);\boldsymbol{\theta}_t\} &= \lim_{dt\to 0}\frac{\Pr\{t\leq T_i^\star<t+dt, K_i=k|T_i^\star\geq t, M_i(t);\boldsymbol{\theta}_t\}}{dt}\\ &= \alpha_{0k}\{t;\boldsymbol{\psi}\}\exp\{\boldsymbol{\gamma}_k^\top\boldsymbol{w}_{ik}+\alpha_k m_i(t)\} \qquad (2.26)\end{aligned}$$

where K_i denotes the failure cause, $\alpha_{ik}\{t|m_i(t);\boldsymbol{\theta}\}$ denotes the cause-specific hazard for cause k, and $\boldsymbol{w}_{ik}$ denotes some baseline covariates associated with cause k. This further implies that the overall survival function is equal to

$$S_i\{t|M_i(t);\boldsymbol{\theta}_t\}=\exp\left[-\sum_{k=1}^{K}\int_0^t\alpha_{0k}\{s;\boldsymbol{\psi}\}\exp\{\boldsymbol{\gamma}_k^\top\boldsymbol{w}_{ik}+\alpha_k m_i(s)\}ds\right], \qquad (2.27)$$

assuming that there are K failure causes. As mentioned in Rizopoulos (2012b) and Bakoyannis and Touloumi (2012), calculating the likelihood when using cause-specific hazards can be easily carried out by transforming the data to the competing-risk long format using K rows for each individual, one for each possible cause. Thus, in general, cause-specific hazard models are computationally easy to implement, which might explain why most competing-risks SREMs are based on cause specific hazards. An exception is the proportional subdistribution hazards joint model proposed by Deslandes

and Chevret (2010), which is defined as

$$\begin{aligned}\lambda_{ik}\{t|m_i(t);\boldsymbol{\theta}_t\} &= \lim_{dt\to 0}\frac{\Pr\{t\leq T_i^\star < t+dt, K_i=k|T_i^\star\geq t\cup(T_i^\star\leq t\cap K_i\neq k), m_i(t);\boldsymbol{\theta}_t\}}{dt}\\ &= \lambda_{0k}\{t;\boldsymbol{\psi}\}\exp\{\boldsymbol{\gamma}_k^\top\boldsymbol{w}_{ik}+\alpha_k m_i(t)\},\end{aligned} \tag{2.28}$$

where $\lambda_{ik}\{t|m_i(t);\boldsymbol{\theta}_t\}$ denotes the subdistribution hazard for cause k. The rationale for assuming models based on the subdistribution hazards is mainly that there is an one-to-one relationship between the CIF for cause k and the corresponding subdistribution hazard, i.e.

$$\lambda_{ik}\{t|m_i(t);\boldsymbol{\theta}_t\} = -\frac{\partial\log\left[1-F_{ik}\{t|M_i(t),\boldsymbol{w}_{ik};\boldsymbol{\theta}_t\}\right]}{\partial t}, \tag{2.29}$$

where $F_{ik}\{t|M_i(t),\boldsymbol{w}_{ik};\boldsymbol{\theta}_t\} = \Pr\{T_i^\star\leq t, K_i=k|M_i(t),\boldsymbol{w}_{ik};\boldsymbol{\theta}_{tk}\}$ denotes the CIF for cause k. In theory, a cause-specific CIF can be obtained from cause-specific hazards by integrating the product of the respective cause-specific hazard and the overall survival function over time, $F_{ik}\{t|M_i(t),\boldsymbol{w}_{ik};\boldsymbol{\theta}_t\} = \int_0^t \alpha_{ik}\{s|m_i(s);\boldsymbol{\theta}_t\}S_i\{s|M_i(s);\boldsymbol{\theta}_t\}ds$ (e.g. Bakoyannis and Touloumi, 2012). On top of that, though, in many SREMs, further complexity emerges as (a) the overall survival function is approximated by numerical integration too (Andrinopoulou et al., 2014, 2017, Rizopoulos, 2012b) or is semiparametrically estimated as a step function (Elashoff et al., 2008, Williamson et al., 2008) and (b) when focusing on dynamic predictions or population-averaged estimates, multidimensional integration over the random effects is also required, e.g. (Andrinopoulou et al., 2017) and (Rizopoulos, 2012b, Chapter 7). Thus, estimating CIFs based on a cause-specific hazards SREM requires complex integration, which is challenging. Therefore, if one focuses on CIFs, an SREM in terms of the CIFs (or some function of them) would be more natural and could substantially reduce the computational burden of formally deriving CIF estimates based on cause-specific hazard estimates.

In standard survival analysis, the literature on regression modelling of CIFs includes numerous approaches, with the most frequently applied method probably being the proportional subdistribution hazards model by Fine and Gray (1999). Ever since, various extensions have been proposed in the literature, including, for example, semiparametric approaches (Bakoyannis et al., 2017, Mao and Lin, 2017), flexibly parametric

approaches (Mozumder et al., 2018), and fully parametric methods based on the Gompertz distribution (Jeong and Fine, 2006). One issue when specifying models for CIFs is that the all-cause CIF should be bounded by 1. In some approaches, this constraint is not formally taken into account (Fine and Gray, 1999, Jeong and Fine, 2006, Mozumder et al., 2018). To some extent, though, an all-cause CIF greater than 1 may indicate model misspecification and can be dealt with by adjusting for all appropriate covariates including also any potential interaction. Shi et al. (2013) tackled this problem by modelling the baseline asymptote for one cause-specific CIF, with the remaining CIFs expressed as a function of this plateau. Mao and Lin (2017) imposed a small positive number on the survival function as a "buffer" to force it to be strictly positive, whereas Bakoyannis et al. (2017) incorporated a formal boundedness constraint during the maximisation process. However, how to impose such a constraint in SREMs may not be so clear as these models are defined conditionally on the random effects, with integration over the prior distribution of the random effects required to obtain the observed data likelihood. Under the Bayesian paradigm, Gelfand et al. (1992) suggested that when the constraints somehow involve the data (as it is the case in CIF modelling), it is more natural to think of the constraints as built into the likelihood function rather than the prior distribution. It was also pointed out that the posterior distribution has the same functional form as it would have had if the constraints had been ignored but it is equal to zero when the constrains are violated (Gelfand et al., 1992). It should be noted, though, that this constraint is automatically fulfilled when applying competing-risks models through cause-specific hazards.

In competing-risk settings, a frequent issue is that the cause of failure is missing or potentially misclassified. For example, in HIV cohort studies, especially in those from resource-constrained countries, significant under-reporting of deaths is a frequent major issue. That is, patients who have actually died may have been incorrectly classified as disengaged from care, which is an example of failure cause misclassification. This implies that there may be serious underestimation of mortality and overestimation of the risk for disengagement from care. To deal with this issue, a standard approach is to

collect a small random sample from the patients reported to be disengaged from care and actively ascertain their true vital status, which is an example of a double sampling design (Cook and Kosorok, 2004). Since this is performed for typically 10-20% of the patients due to financial constraints, the "true" failure causes for the remaining patients are missing. Various methods to adjust for outcome misclassification using double sampling data have been proposed in the literature. For example, Bakoyannis et al. (2019) used a pseudo-likelihood approach to account for missing absorbing states in a multistate model, Daniel Paulino et al. (2003) accounted for outcome misclassification in Binomial regression using MCMC, and Lyles et al. (2011) used a direct maximum likelihood approach in logistic regression under outcome misclassification.

Another setting where a competing-risk SREM could be particularly useful is to describe the progression of a study population in terms of multistate probabilities based on marker and competing-risk data. For example, in applied epidemiological/medical research, especially for prediction purposes, progression of cohorts over time may be monitored by using mutually exclusive states defined jointly by competing-risk data and discretised continuous marker data. For example, the United Nations (UN) Joint Programme in HIV/AIDS (UNAIDS) produces various estimates of parameters relevant to the worldwide HIV epidemic using the following CD4 states: [0,50), [50,100), [100,200), [200,250), [250,350), [350,500), [500,∞] cells/μL, through the Spectrum software (Stover et al., 2019). In Spectrum, the CD4 states are considered along with many clinical events/endpoints, e.g. death, onset of AIDS, and disengagement from care. In chronic kidney disease longitudinal studies, the Glomerular filtration rate (GFR) is a key surrogate of kidney function, which is typically collected over time. Survival states (death and end-stage kidney disease) together with discrete states based on the GFR levels have been used in this context (Hu et al., 2012). Hu et al. (2012) proposed an interesting approach estimating multistate probabilities defined jointly by unobserved marker data and observed competing-risk data. However, estimates are not based on a single unified model but rather they are based on a two-stage approach which does not account for potential failure cause misclassification. At the first stage, the un-

observed marker data are estimated by subject-specific regression models and, at the second stage, marker states are obtained by averaging over individuals accounting for the randomness of the estimated marker trajectories. Thus, the effects of the true marker values on the competing risks are not explicit and it is questionable how well the method would perform under highly irregular visit times and a considerable proportion of individuals with one marker measurement (as it is the case in CD4 modeling and many other examples).

Pattern-mixture models

In pattern-mixture models, a different factorization of the joint distribution of $(\boldsymbol{Y}_i, M_i)$ is used. Specifically, one assumes a model for the conditional distribution of the marker given the dropout pattern and then a model for the marginal probability of dropout, i.e. $f(\boldsymbol{Y}_i, M_i; \boldsymbol{\theta}_L, \boldsymbol{\theta}_t) = f(\boldsymbol{Y}_i | M_i; \boldsymbol{\theta}_L) \Pr(M_i; \boldsymbol{\theta}_t)$. Based on this factorization, it is clear that the conditional distribution of the marker given the dropout pattern is not entirely identifiable for all patterns except for the "completers" as marker values after dropout are missing. Thus, restrictions must be built into the model to ensure that there are links among the distributions of the outcomes conditional on the patterns of non-response (Little, 1993, 1994, Little and Wang, 1996). The marker probabilities of dropout, $\Pr(M_i = j; \boldsymbol{\theta}_t), j = 1, 2, \ldots, Q$, can definitely be estimated using the observed data, where Q denotes the number of dropout patterns in the data. To estimate the marginal marker distribution, which is often of primary interest, one can apply the formula $f(\boldsymbol{Y}_i; \boldsymbol{\theta}_L) = \sum_{j=1}^{Q} f(\boldsymbol{Y}_i | M_i = j; \boldsymbol{\theta}_L) \Pr(M_i = j; \boldsymbol{\theta}_t)$.

Chapter 3

Performance of longitudinal and time-to-dropout joint models when the dropout mechanism is at random

3.1 Introduction

In the previous chapter, we presented an overview of various methods for modeling a longitudinal marker when there are incomplete data, juxtaposing the different assumptions of the available approaches. Most methods make, implicitly or explicitly, assumptions about the joint distribution of the marker and the dropout processes. The take-home message is that the validity of results from a model applied to incomplete data depends crucially on the nature of the true missing data mechanism, i.e. on how the probabilities of missingness depend on the full set of marker values.

In the missing data literature, it has long been emphasized that when there are missing data, it is not always possible to verify the assumptions about the missing data mechanisms based on the observed data. In the joint modeling literature, though,

especially in SREMs, the consequences of this result are sometimes ignored. Thus, in practice, when the nature of the dropout mechanism is unknown, many researchers may apply SREMs, at least as part of a set of sensitivity analyses, if informative dropout is suspected. Following standard recommendations from published papers (e.g. Gras et al., 2013, Schluchter, 1992, Touloumi et al., 1999), researchers may, perhaps wrongly, interpret deviations between results from LMMs (assuming ignorable dropout) and from SREMs as indicating MNAR missingness. The discrimination though between MAR and MNAR is not possible in practice, as it depends on untestable assumptions (Molenberghs et al., 2008). Thus one may end up fitting an SREM to data that are actually MAR. However, even though the bias of LMMs when fitted to MNAR data has been thoroughly investigated, what has not been fully studied is the performance of SREM models when the true dropout mechanism is MAR.

The motivation for this research question has been clearly motivated by issues arising in the CD4 cell count modeling during the HIV natural history. In the earlier years of the HIV epidemic, censoring of CD4 counts was mainly due to AIDS onset or death. This is considered to correspond to an MNAR mechanism, and thus several SREMs have been applied to adjust for that. However, nowadays, the primary source of censoring is treatment initiation, with the nature of this mechanism being still debatable. Some authors argue that it should be considered to be MAR as cART is initiated mainly based on observed values of CD4 counts (Pantazis et al., 2005); according to guidelines up to 2015, cART should be initiated when CD4 counts drop below a certain level. On the other hand, other researchers suggest adjusting for informative dropout due to cART initiation, as results from joint models were substantially different from those from LMMs (Gras et al., 2013). This issue is further complicated by the fact that cART initiation was also recommended when very high (i.e. $> 10^5$ copies/mL) viral load (VL) levels are observed. VL is another marker of disease progression quantifying the amount of virus in blood, which is also measured longitudinally over time. VL and CD4 are correlated and they may be both related to dropout probabilities; thus VL may act as an auxiliary variable which is required in order to make the MAR assump-

tion more reliable. If so, a bivariate model, modeling simultaneously both markers may be necessary.

In this chapter, we examine the case in which an SREM model is applied to incomplete MAR data. In Section 3.2 we derive the bias in the slope estimate from the LN-SREM(RE) model under two specific MAR dropout mechanisms. The corresponding bias in the PH-SREM(CV) model is derived in Section 3.3. A proposed alternative model, along with the fitting procedure, are described in Section 3.4. The performance of the proposed model is evaluated in a simulation study in Section 3.5. In Section 3.6 the proposed model, as well as other SREMs, are fitted to real data from the "Concerted action on seroconversion to AIDS and death in Europe" (CASCADE) study, a collaboration of cohorts of HIV-infected individuals with known seroconversion dates (Pantazis et al., 2016). Section 3.7 presents concluding remarks and discusses possible extensions.

3.2 Asymptotic Bias in the LN-SREM(RE) Model when Fitted to MAR Data

Let $\boldsymbol{Y}^\top = (Y_1, \ldots, Y_Q)$ be the measurements of the response variable (a disease marker, e.g. CD4 count), intended to be collected at the same times $t_1, \ldots, t_Q$ on a randomly chosen subject i. Index i has been suppressed for simplicity. Let us also assume that missingness arises only from subjects' dropout. Recall that since missingness is assumed to be monotone, the missingness indicators $\boldsymbol{R}^\top = (R_1, \ldots, R_Q)$, where $R_j = 1$ if Y_j is observed and $R_j = 0$ otherwise, can be replaced by $M = \sum_{j=1}^{Q} R_j$, the number of the observed marker values. Supposing that $M = j$, the full data can be factorized as $\boldsymbol{Y}^\top = (\boldsymbol{Y}_{(j)}^\top, \boldsymbol{Y}_{(\bar{j})}^\top)$, where $\boldsymbol{Y}_{(j)}^\top = (Y_1, \ldots, Y_j)$ is the observed part of $\boldsymbol{Y}$ and $\boldsymbol{Y}_{(\bar{j})}^\top = (Y_{j+1}, \ldots, Y_Q)$ is the missing part of $\boldsymbol{Y}$. $M = Q$ implies that the full data are observed. We assume that missingness is MAR, i.e. the dropout probabilities, $\Pr(M = j|\boldsymbol{Y}_{(j)}; \boldsymbol{\theta}_{t0})$, depend only on the observed data for each dropout pattern and every realization of the observed data $\boldsymbol{Y}_{(j)}, j = 1, \ldots, Q$ (Seaman et al., 2013). Let

T be the time to dropout and $\delta = I(M < Q)$ be the dropout indicator. We assume that dropout happens 0.1 years after the last observed measurement, i.e. $T = t_j + k$ for $j < Q$ with $k = 0.1$ years, whereas $T = t_Q$ for $M = Q$. We set $k = 0.1$ as in our motivating example, when patients fulfill the criteria for treatment initiation it typically takes approximately one month on average to start antiretroviral treatment.

We assume that the true marker evolution is based on an LMM of the form $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta}_0 + \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{\epsilon}$; recall that $\boldsymbol{X}$ and $\boldsymbol{Z}$ denote the design matrices for the fixed and random effects at the scheduled times, respectively, $\boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \sigma_0^2 \boldsymbol{I}_Q)$ are the within-subject residuals and $\boldsymbol{b} \sim N(\boldsymbol{0}, \sigma_0^2 \boldsymbol{D}_0)$ are the random effects. Thus, the true mean and covariance matrix of $\boldsymbol{Y}$ are equal to $E_0(\boldsymbol{Y}) = \boldsymbol{X}\boldsymbol{\beta}_0$ and $Var_0(\boldsymbol{Y}) = \sigma_0^2(\boldsymbol{I} + \boldsymbol{Z}\boldsymbol{D}_0\boldsymbol{Z}^\top)$, respectively. Subscript "0" indicates the true parameter values while the notation E_0 and Var_0 is used to emphasize that the expectations are taken with respect to the true data generating mechanism. For simplicity, no covariates will be assumed, thus the matrices $\boldsymbol{X}$ and $\boldsymbol{Z}$ both include a column of 1's and a column with the measurement times. In this setting, $\boldsymbol{\beta}_0$ includes the true values of the population average at time zero and the time-slope. Our interest lies on the second element of $\boldsymbol{\beta}_0$, i.e. the time-slope, which is our target estimand throughout this part of the thesis.

Recall that in the LN-SREM(RE) model, the marker is modeled through an LMM as described above, whereas the log-time to dropout, U, is described through a model of the form: $U = \zeta + \boldsymbol{\lambda}^\top \boldsymbol{b} + r$, with $r \sim N(0, \nu^2\sigma^2)$, ζ being the expected log-time to dropout and $\nu^2\sigma^2$ being the variance of the log time to dropout. Note that $r \sim N(0, \nu^2\sigma^2)$ corresponds to a re-parameterization of the variance of the log-time to dropout which does not affect the model's estimates. In this formulation, the key parameter is $\boldsymbol{\lambda}$, the parameter measuring the association between the two processes. In our special case, $Y_j = (\beta_0 + b_0) + (\beta_1 + b_1)t_j + \epsilon_j$, $j = 1, \ldots, Q$, and $U = \zeta + \lambda_0 b_0 + \lambda_1 b_1 + r$, where λ_0 and λ_1 relate the dropout time to the subject-specific baseline marker values and marker rate of change (slope), respectively. That is, positive values of λ_1 imply that subjects with slopes lower (i.e. steeper) than the population average β_1 (i.e. $b_1 < 0$) are more likely to drop out of the study earlier. If the association parameters are non-zero,

the model indicates MNAR mechanism, conditional on the model's assumptions being correct (Gruttola and Tu, 1994, Touloumi et al., 1999).

Given $M = j$ for $j < Q$, the likelihood contribution is $f(\boldsymbol{Y}_{(j)};\boldsymbol{\theta}_L)f(u_j|\boldsymbol{Y}_{(j)};\boldsymbol{\theta})$, where $\boldsymbol{\theta}^\top = (\boldsymbol{\beta}^\top, \boldsymbol{\lambda}^\top, \text{vech}(\boldsymbol{D})^\top, \zeta, \nu, \sigma^2)$ is the parameter vector of the LN-SREM(RE) model and $u_j = \log(t_j+k)$ the assumed log-time to dropout; let $\boldsymbol{\theta}_L^\top = (\boldsymbol{\beta}^\top, \text{vech}(\boldsymbol{D})^\top, \sigma^2)$ denote the corresponding parameters of the marker model. Using standard results, the mean and covariance matrix of $U|\boldsymbol{Y}_{(j)};\boldsymbol{\theta}$ are equal to

$$\begin{aligned} E(U|\boldsymbol{Y}_{(j)};\boldsymbol{\theta}) &= \zeta + \boldsymbol{\lambda}^\top \boldsymbol{D}\boldsymbol{Z}_{(j)}^\top \boldsymbol{V}_{(j)}^{-1}(\boldsymbol{Y}_{(j)} - \boldsymbol{X}_{(j)}\boldsymbol{\beta}) \\ Var(U|\boldsymbol{Y}_{(j)};\boldsymbol{\theta}) &= \sigma^2(\nu^2 + \boldsymbol{\lambda}^\top \boldsymbol{D}\boldsymbol{\lambda} - \boldsymbol{\lambda}^\top \boldsymbol{D}\boldsymbol{Z}_{(j)}^\top \boldsymbol{V}_{(j)}^{-1}\boldsymbol{Z}_{(j)}\boldsymbol{D}\boldsymbol{\lambda}), \end{aligned}$$

respectively, where $\boldsymbol{X}_{(j)}$ and $\boldsymbol{Z}_{(j)}$ are the appropriate submatrices of $\boldsymbol{X}$ and $\boldsymbol{Z}$, and $\boldsymbol{V}_{(j)} = (\boldsymbol{I}_{(j)} + \boldsymbol{Z}_{(j)}\boldsymbol{D}\boldsymbol{Z}_{(j)}^\top)$. When $M = Q$, $f(u_Q|\boldsymbol{Y}_{(Q)};\boldsymbol{\theta})$ has to be replaced by $\int_{u_Q}^{+\infty} f(x|\boldsymbol{Y}_{(Q)};\boldsymbol{\theta})dx$, where $u_Q = \log(t_Q)$. Thus, the likelihood function for a randomly sampled individual from the population is equal to

$$f(\boldsymbol{Y}_{(M)}, U;\boldsymbol{\theta}) = \begin{cases} f(\boldsymbol{Y}_{(1)};\boldsymbol{\theta}_L)f(u_1|\boldsymbol{Y}_{(1)};\boldsymbol{\theta}) & \text{if } M = 1; \\ f(\boldsymbol{Y}_{(2)};\boldsymbol{\theta}_L)f(u_2|\boldsymbol{Y}_{(2)};\boldsymbol{\theta}) & \text{if } M = 2; \\ \vdots & \\ f(\boldsymbol{Y}_{(Q)};\boldsymbol{\theta}_L)\int_{u_Q}^{\infty} f(x|\boldsymbol{Y}_{(Q)};\boldsymbol{\theta})dx & \text{if } M = Q. \end{cases}$$

To derive the asymptotic bias of the LN-SREM(RE) slope estimator, we use a method identifying the limit in probability of estimators under incomplete data (e.g. Manski, 1988, Newey and McFadden, 1994, Rotnitzky and Wypij, 1994). This requires solving $E_0\{\mathcal{U}(\boldsymbol{\theta})\} = \boldsymbol{0}$, where $\mathcal{U}(\boldsymbol{\theta}) = \frac{\partial}{\partial\boldsymbol{\theta}}\log f(\boldsymbol{Y}_{(M)}, U;\boldsymbol{\theta})$ is the score vector of the LN-SREM(RE) model, with the expectation taken over the true distribution of the data and the dropout process, evaluated at the true parameter values, i.e. E_0 is a shorthand for $E_{(\boldsymbol{Y}_{(M)},M)}\{\cdot;\boldsymbol{\theta}_0\}$, with $\boldsymbol{\theta}_0$ being the true parameter value of the distribution of

$(\boldsymbol{Y}_{(M)}, M)$. Therefore, we need to solve the following system of equations

$$\begin{aligned} & \sum_{j=1}^{Q} \int \frac{\partial}{\partial \boldsymbol{\theta}} \left\{ \log f(\boldsymbol{Y}_{(j)}; \boldsymbol{\theta}_L) \right\} f_0(\boldsymbol{Y}_{(j)}, M = j; \boldsymbol{\theta}_0) d\boldsymbol{Y}_{(j)} \\ + & \sum_{j=1}^{Q-1} \int \frac{\partial}{\partial \boldsymbol{\theta}} \left\{ \log f(u_j | \boldsymbol{Y}_{(j)}; \boldsymbol{\theta}) \right\} f_0(\boldsymbol{Y}_{(j)}, M = j; \boldsymbol{\theta}_0) d\boldsymbol{Y}_{(j)} \\ + & \int \frac{\partial}{\partial \boldsymbol{\theta}} \left\{ \log \int_{u_Q}^{\infty} f(x | \boldsymbol{Y}_{(Q)}; \boldsymbol{\theta}) dx \right\} f_0(\boldsymbol{Y}_{(Q)}, M = Q; \boldsymbol{\theta}_0) d\boldsymbol{Y}_{(Q)} = \mathbf{0}, \end{aligned} \quad (3.1)$$

where $f_0(\boldsymbol{Y}_{(j)}, M = j; \boldsymbol{\theta}_0)$ denotes the likelihood of the "true" joint distribution of the marker and the dropout process, evaluated at the true parameter values. Interchanging the order of differentiation and integration, we have that the first term in Equation (3.1) (i.e. the contribution of the marker model) equals

$$\sum_{j=1}^{Q} \frac{\partial}{\partial \boldsymbol{\theta}} a_{0j} E_0 \left\{ -\frac{j \log(\sigma^2)}{2} - \frac{\log |\boldsymbol{V}_{(j)}|}{2} - \frac{(\boldsymbol{Y}_{(j)} - \boldsymbol{X}_{(j)}\boldsymbol{\beta})^\top \boldsymbol{V}_{(j)}^{-1} (\boldsymbol{Y}_{(j)} - \boldsymbol{X}_{(j)}\boldsymbol{\beta})}{2\sigma^2} \middle| M = j; \boldsymbol{\theta}_0 \right\},$$

where $a_{0j} = \Pr(M = j; \boldsymbol{\theta}_0)$ denotes the true probability of the jth dropout pattern. By the formula $E\left\{(\boldsymbol{Y} - \boldsymbol{a})^\top \boldsymbol{W} (\boldsymbol{Y} - \boldsymbol{a})\right\} = (\boldsymbol{\mu} - \boldsymbol{a})^\top \boldsymbol{W} (\boldsymbol{\mu} - \boldsymbol{a}) + \mathrm{tr}(\boldsymbol{\Omega}\boldsymbol{W})$, where $\boldsymbol{Y}$ is any random vector with $E(\boldsymbol{Y}) = \boldsymbol{\mu}$ and $Var(\boldsymbol{Y}) = \boldsymbol{\Omega}$, it follows that the contribution of the marker model equals

$$\begin{aligned} & \sum_{j=1}^{Q} \frac{\partial}{\partial \boldsymbol{\theta}} \Big\{ -\frac{j a_{0j} \log(\sigma^2)}{2} - \frac{a_{0j} \log |\boldsymbol{V}_{(j)}|}{2} \\ & - \frac{a_{0j} (\boldsymbol{\mu}_{0(j)} - \boldsymbol{X}_{(j)}\boldsymbol{\beta})^\top \boldsymbol{V}_{(j)}^{-1} (\boldsymbol{\mu}_{0(j)} - \boldsymbol{X}_{(j)}\boldsymbol{\beta}) + a_{0j} \mathrm{tr}(\boldsymbol{V}_{(j)}^{-1} \boldsymbol{\Omega}_{0(j)})}{2\sigma^2} \Big\}, \end{aligned} \quad (3.2)$$

where $\boldsymbol{\mu}_{0(j)}$ and $\boldsymbol{\Omega}_{0(j)}$ denote the mean and covariance matrix of the true distribution of $\boldsymbol{Y}_{(j)} | M = j; \boldsymbol{\theta}_0$, respectively. After some straightforward algebra, the log-density of the dropout model for the jth dropout pattern ($j < Q$) can be shown to be equal to

$$\begin{aligned} - & \frac{\log Var(U | \boldsymbol{Y}_{(j)}; \boldsymbol{\theta})}{2} - \frac{1}{2Var(U | \boldsymbol{Y}_{(j)}; \boldsymbol{\theta})} \Big[u_j^2 - 2u_j \left\{ \zeta + \boldsymbol{\omega}_{(j)}^\top (\boldsymbol{Y}_{(j)} - \boldsymbol{X}_{(j)}\boldsymbol{\beta}) \right\} \\ + & \zeta^2 + 2\zeta \boldsymbol{\omega}_{(j)}^\top (\boldsymbol{Y}_{(j)} - \boldsymbol{X}_{(j)}\boldsymbol{\beta}) + (\boldsymbol{Y}_{(j)} - \boldsymbol{X}_{(j)}\boldsymbol{\beta})^\top \boldsymbol{A}_{(j)} (\boldsymbol{Y}_{(j)} - \boldsymbol{X}_{(j)}\boldsymbol{\beta}) \Big], \end{aligned} \quad (3.3)$$

with $\boldsymbol{\omega}_{(j)}^{\top} = \boldsymbol{\lambda}^{\top}\boldsymbol{D}\boldsymbol{Z}_{(j)}^{\top}\boldsymbol{V}_{(j)}^{-1}$ and $\boldsymbol{A}_{(j)} = \boldsymbol{\omega}_{(j)}\boldsymbol{\omega}_{(j)}^{\top}$. By taking the expectation of Equation (3.3) over the true distribution of $\boldsymbol{Y}_{(j)}|M = j;\boldsymbol{\theta}_0$ and observing that $Var(U|\boldsymbol{Y}_{(j)};\boldsymbol{\theta})$ and $\boldsymbol{\omega}_{(j)}$ do not depend on $\boldsymbol{Y}_{(j)}$ (thus $Var(U|\boldsymbol{Y}_{(j)};\boldsymbol{\theta}) = Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})$), we arrive at

$$
\begin{aligned}
- \quad & \frac{\log Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{2} - \frac{1}{2Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}\Bigg[u_j^2 - 2u_j\left\{\zeta + \boldsymbol{\omega}_{(j)}^{\top}(\boldsymbol{\mu}_{0(j)} - \boldsymbol{X}_{(j)}\boldsymbol{\beta})\right\} \\
+ \quad & \zeta^2 + 2\zeta\boldsymbol{\omega}_{(j)}^{\top}(\boldsymbol{\mu}_{0(j)} - \boldsymbol{X}_{(j)}\boldsymbol{\beta}) + (\boldsymbol{\mu}_{0(j)} - \boldsymbol{X}_{(j)}\boldsymbol{\beta})^{\top}\boldsymbol{A}_{(j)}(\boldsymbol{\mu}_{0(j)} - \boldsymbol{X}_{(j)}\boldsymbol{\beta}) + \mathrm{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)})\Bigg].
\end{aligned}
$$

Next, letting $E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta}) = \zeta + \boldsymbol{\lambda}^{\top}\boldsymbol{D}\boldsymbol{Z}_{(j)}^{\top}\boldsymbol{V}_{(j)}^{-1}(\boldsymbol{\mu}_{0(j)} - \boldsymbol{X}_{(j)}\boldsymbol{\beta})$ be the mean of the log-time to dropout evaluated at the true mean of $\boldsymbol{Y}_{(j)}|M = j;\boldsymbol{\theta}$, it can be easily shown by completing the square that the second row in Equation (3.1) is equal to

$$
\begin{aligned}
\sum_{j=1}^{Q-1}\frac{\partial}{\partial\boldsymbol{\theta}}\Bigg[& -\frac{a_{0j}\log Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{2} - \frac{a_{0j}}{2Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}\left\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\right\}^2 \\
& - \frac{a_{0j}}{2Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}\mathrm{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)})\Bigg]. \quad (3.4)
\end{aligned}
$$

Therefore, replacing the first two lines of Equation (3.1) by Equation (3.2) and Equation (3.4), respectively, leads to

$$
\begin{aligned}
& \frac{\partial}{\partial\boldsymbol{\theta}}\Bigg(\sum_{j=1}^{Q}\Bigg[\frac{-ja_{0j}\log(\sigma^2)}{2} - \frac{a_{0j}\log|\boldsymbol{V}_{(j)}|}{2} \\
& - \frac{a_{0j}\left\{(\boldsymbol{\mu}_{0(j)} - \boldsymbol{X}_{(j)}\boldsymbol{\beta})^{\top}\boldsymbol{V}_{(j)}^{-1}(\boldsymbol{\mu}_{0(j)} - \boldsymbol{X}_{(j)}\boldsymbol{\beta}) + \mathrm{tr}(\boldsymbol{V}_{(j)}^{-1}\boldsymbol{\Omega}_{0(j)})\right\}}{2\sigma^2}\Bigg] + \\
& \sum_{j=1}^{Q-1}\Bigg[-\frac{a_{0j}\log Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{2} - \frac{a_{0j}\left\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\right\}^2 + a_{0j}\mathrm{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)})}{2Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}\Bigg] \\
& + a_{0Q}E_0\left\{\log\int_{u_Q}^{\infty} f(x|\boldsymbol{Y}_{(Q)};\boldsymbol{\theta})dx\,\middle|\,M = Q;\boldsymbol{\theta}_0\right\}\Bigg) = \boldsymbol{0}. \quad (3.5)
\end{aligned}
$$

To solve Equation (3.5), one could use the the Newton-Raphson algorithm. This, however, requires, the derivative of (3.5) with respect to $\boldsymbol{\theta}$, which are presented in detail in the following subsection.

3.2.1 Derivatives of Equation (3.5) over $\boldsymbol{\theta}$

We now present the first- and second-order partial derivatives of Equation (3.5), except for the pattern of no dropout for which the central difference approximation was used (Press et al., 2007). These matrix differentiation formulas were mainly based on results from Harville (1997) and Lindstrom and Bates (1988). To save space we show results from the jth summand of (2). Using straightforward matrix differentiation rules, the first-order partial derivatives with respect to $\boldsymbol{\lambda}$ are

$$\mathcal{U}_j(\boldsymbol{\lambda}) = -\frac{a_{0j}\frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\boldsymbol{\lambda}}}{2Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})} - \frac{a_{0j}}{2Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}\left[\frac{\partial\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2}{\partial\boldsymbol{\lambda}} + \frac{\partial \mathrm{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)})}{\partial\boldsymbol{\lambda}}\right]$$
$$+ \frac{a_{0j}}{2}\left[\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2 + \mathrm{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)})\right]\frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\boldsymbol{\lambda}}/Var^2(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta}), \tag{3.6}$$

where $\frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\boldsymbol{\lambda}} = 2\sigma^2(\boldsymbol{D} - \boldsymbol{D}\boldsymbol{Z}_{(j)}^\top\boldsymbol{V}_{(j)}^{-1}\boldsymbol{Z}_{(j)}\boldsymbol{D})\boldsymbol{\lambda}$, $\frac{\partial\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2}{\partial\boldsymbol{\lambda}} = 2\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}\frac{\partial\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}}{\partial\boldsymbol{\lambda}}$ with $\frac{\partial\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}}{\partial\boldsymbol{\lambda}} = -\boldsymbol{D}\boldsymbol{Z}_{(j)}^\top\boldsymbol{V}_{(j)}^{-1}\boldsymbol{r}_{(j)}$, and $\frac{\partial \mathrm{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)})}{\partial\boldsymbol{\lambda}} = 2\boldsymbol{D}\boldsymbol{Z}_{(j)}^\top\boldsymbol{V}_{(j)}^{-1}\boldsymbol{\Omega}_{0(j)}\boldsymbol{V}_{(j)}^{-1}\boldsymbol{Z}_{(j)}\boldsymbol{D}\boldsymbol{\lambda}$. Note that $\mathrm{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)})$ can be written equivalently as $\boldsymbol{\omega}_{(j)}^\top\boldsymbol{A}_{(j)}\boldsymbol{\omega}_{(j)}$, by the basic properties of the trace operator.

Similarly, the partial derivatives with respect to $\boldsymbol{\beta}$ are equal to

$$\mathcal{U}_j(\boldsymbol{\beta}) = \frac{a_{0j}}{\sigma^2}\boldsymbol{X}_{(j)}^\top\boldsymbol{V}_{(j)}^{-1}\boldsymbol{r}_{(j)} - \frac{a_{0j}}{2Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}\frac{\partial\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2}{\partial\boldsymbol{\beta}}, \tag{3.7}$$

where $\frac{\partial\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2}{\partial\boldsymbol{\beta}} = 2\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}\frac{\partial\left\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\right\}}{\partial\boldsymbol{\beta}}$ and $\frac{\partial\left\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\right\}}{\partial\boldsymbol{\beta}} = \boldsymbol{X}_{(j)}^\top\boldsymbol{V}_{(j)}^{-1}\boldsymbol{Z}_{(j)}\boldsymbol{D}\boldsymbol{\lambda}$. Also, the partial derivative with respect to ζ is just $a_{0j}\left\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\right\}/Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})$, whereas the first-order derivative with respect to $\phi = \log(\nu^2)$ is

$$\mathcal{U}_j(\phi) = -\frac{a_{0j}\frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\phi}}{2Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})} + \frac{a_{0j}\frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\phi}}{2Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}\left[\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2 + \mathrm{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)})\right], \tag{3.8}$$

with $\frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\phi} = \sigma^2 e^\phi$. Moving on the calculations regarding the Hessian matrix,

we first give the $\boldsymbol{\lambda}$ block of it, which is equal to

$$\begin{aligned}\mathcal{H}_j(\boldsymbol{\lambda}) &= -\frac{a_{0j}}{2Var^2(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}\left\{Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\frac{\partial^2 Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}^\top}\right. \\ &- \left.\frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\boldsymbol{\lambda}}\frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\boldsymbol{\lambda}^\top}\right\} - \frac{a_{0j}}{2}\frac{\partial}{\partial\boldsymbol{\lambda}^\top}\left\{\frac{f_j(\boldsymbol{\lambda})}{g_j(\boldsymbol{\lambda})}\right\},\end{aligned} \tag{3.9}$$

where $\frac{\partial^2 Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}^\top} = 2\sigma^2(\boldsymbol{D} - \boldsymbol{D}\boldsymbol{Z}_{(j)}^\top\boldsymbol{V}_{(j)}^{-1}\boldsymbol{Z}_{(j)}\boldsymbol{D})$, with

$$\begin{aligned}f_j(\boldsymbol{\lambda}) &= Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\left[\frac{\partial\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2}{\partial\boldsymbol{\lambda}} + \frac{\partial\text{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)})}{\partial\boldsymbol{\lambda}}\right] \\ &- \left[\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2 + \text{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)})\right]\frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\boldsymbol{\lambda}}\end{aligned}$$

and $g_j(\boldsymbol{\lambda}) = Var^2(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})$ defined to facilitate the presentation of the results. Then

$$\begin{aligned}\frac{\partial}{\partial\boldsymbol{\lambda}^\top}f_j(\boldsymbol{\lambda}) &= \left[\frac{\partial\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2}{\partial\boldsymbol{\lambda}} + \frac{\partial\text{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)})}{\partial\boldsymbol{\lambda}}\right]\frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\boldsymbol{\lambda}^\top} \\ &+ Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\left[\frac{\partial^2\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}^\top} + \frac{\partial^2\text{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)})}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}^\top}\right] \\ &- \frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\boldsymbol{\lambda}}\left[\frac{\partial\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2}{\partial\boldsymbol{\lambda}^\top} + \frac{\partial\text{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)})}{\partial\boldsymbol{\lambda}^\top}\right] \\ &- \left[\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2 + \text{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)})\right]\frac{\partial^2 Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}^\top},\end{aligned}$$

where

$$\frac{\partial^2\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}^\top} = 2\frac{\partial\left\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\right\}}{\partial\boldsymbol{\lambda}}\frac{\partial\left\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\right\}}{\partial\boldsymbol{\lambda}^\top}$$

and $\frac{\partial^2\text{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)})}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}^\top} = 2\boldsymbol{D}\boldsymbol{Z}_{(j)}^\top\boldsymbol{V}_{(j)}^{-1}\boldsymbol{\Omega}_{0(j)}\boldsymbol{V}_{(j)}^{-1}\boldsymbol{Z}_{(j)}\boldsymbol{D}$. It also follows that $\frac{\partial}{\partial\boldsymbol{\lambda}^\top}g_j(\boldsymbol{\lambda}) = 2Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta}) \times \frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\boldsymbol{\lambda}^\top}$, thus

$$\frac{\partial}{\partial\boldsymbol{\lambda}^\top}\left\{\frac{f_j(\boldsymbol{\lambda})}{g_j(\boldsymbol{\lambda})}\right\} = \frac{g_j(\boldsymbol{\lambda})\frac{\partial}{\partial\boldsymbol{\lambda}^\top}f_j(\boldsymbol{\lambda}) - f_j(\boldsymbol{\lambda})\frac{\partial}{\partial\boldsymbol{\lambda}^\top}g_j(\boldsymbol{\lambda})}{g_j^2(\boldsymbol{\lambda})}$$

completes the calculations required for the $\boldsymbol{\lambda}$-block of the Hessian matrix. Similarly,

the $\boldsymbol{\beta}$-block of the Hessian matrix is equal to

$$\begin{aligned}\mathcal{H}_j(\boldsymbol{\beta}) &= \frac{a_{0j}}{\sigma^2}\boldsymbol{X}_{(j)}^\top \boldsymbol{V}_{(j)}^{-1}\boldsymbol{X}_{(j)} - \frac{a_{0j}}{Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}\frac{\partial\left\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\right\}}{\partial\boldsymbol{\beta}} \\ &\times \frac{\partial\left\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\right\}}{\partial\boldsymbol{\beta}^\top}, \end{aligned} \tag{3.10}$$

whereas, after some straightforward manipulations, the $(\boldsymbol{\beta},\boldsymbol{\lambda})$ block can be shown to be equal to

$$\begin{aligned}\mathcal{H}(\boldsymbol{\beta},\boldsymbol{\lambda}) &= -\frac{a_{0j}}{Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}\left[\frac{\partial\left\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\right\}}{\partial\boldsymbol{\lambda}}\frac{\partial\left\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\right\}}{\partial\boldsymbol{\beta}^\top}\right. \\ &+ \left.\left\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\right\}\frac{\partial^2\left\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\right\}}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\beta}^\top}\right] \\ &+ \frac{a_{0j}}{Var^2(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}\left\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\right\} \\ &\times \frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\boldsymbol{\lambda}}\frac{\partial\left\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\right\}}{\partial\boldsymbol{\beta}^\top}, \end{aligned} \tag{3.11}$$

with $\frac{\partial^2\left\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\right\}}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\beta}^\top} = \boldsymbol{D}\boldsymbol{Z}_{(j)}^\top\boldsymbol{V}_{(j)}^{-1}\boldsymbol{X}_{(j)}$. The ζ-block of the Hessian matrix is just $-a_{0j}/Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})$, whereas, after some algebra, the ϕ block can be written as

$$\begin{aligned}\mathcal{H}_j(\phi) &= -\frac{a_{0j}}{2}\frac{\frac{\partial^2 Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial^2\phi}Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta}) - \left\{\frac{\partial^2 Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\phi}\right\}^2}{Var^2(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})} \\ &+ \frac{a_{0j}}{2}\left[\frac{\frac{\partial^2 Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial^2\phi}Var^2(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta}) - 2Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\left\{\frac{\partial^2 Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\phi}\right\}^2}{Var^4(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}\right] \\ &\times \left[\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2 + \mathrm{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)})\right], \end{aligned} \tag{3.12}$$

with $\frac{\partial^2 Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial^2\phi} = \sigma^2 e^\phi$. The (ϕ,ζ) block simply is

$$\mathcal{H}_j(\phi,\zeta) = -a_{0j}\frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\phi} \times \{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2 \Big/ Var^2(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta}). \tag{3.13}$$

The $(\boldsymbol{\beta},\zeta)$ block is equal to $\mathcal{H}(\boldsymbol{\beta},\zeta) = \frac{a_{0j}}{Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}\frac{\partial\left\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\right\}}{\partial\boldsymbol{\beta}}$ whereas the $(\boldsymbol{\beta},\phi)$ block is equal to $\mathcal{H}(\boldsymbol{\beta},\phi) = a_{0j}\frac{\partial\left\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\right\}^2}{\partial\boldsymbol{\beta}}\frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\phi}/Var^2(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})$. The

$(\boldsymbol{\lambda}, \zeta)$ block is

$$\begin{aligned}\mathcal{H}(\boldsymbol{\lambda}, \zeta) &= -\frac{a_{0j}}{2Var^2(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}\Bigg[-2Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\frac{\partial\left\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\right\}}{\partial\boldsymbol{\lambda}} \\ &+ 2\left\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\right\}\frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\boldsymbol{\lambda}}\Bigg], \qquad (3.14)\end{aligned}$$

and the $(\boldsymbol{\lambda}, \phi)$ block is

$$\begin{aligned}\mathcal{H}(\boldsymbol{\lambda}, \phi) &= +\frac{a_{0j}}{2}\frac{\frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\phi}}{Var^2(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}\frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\boldsymbol{\lambda}} \\ &+ \frac{a_{0j}}{2}\frac{\frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\phi}}{Var^2(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}\left[\frac{\partial\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2}{\partial\boldsymbol{\lambda}} + \frac{\partial\mathrm{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)})}{\partial\boldsymbol{\lambda}}\right] \\ &- \frac{a_{0j}\frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\phi}}{Var^3(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}\frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\boldsymbol{\lambda}} \\ &\times \left[\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2 + \mathrm{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)})\right]. \qquad (3.15)\end{aligned}$$

Next, following the proposal in Lindstrom and Bates (1988), we present the derivatives with respect to the variance components. As shown by Lindstrom and Bates (1988), the Choleski factorization of $\boldsymbol{D}$ (i.e. $\boldsymbol{D} = \boldsymbol{L}^\top\boldsymbol{L}$, with $\boldsymbol{L}$ being an upper-triangular matrix) dramatically improves the converge properties of the Newton-Raphson algorithm. In fact, to facilitate the computations, we first calculated the derivatives over $\mathrm{vec}(\boldsymbol{D})$ assuming that $\boldsymbol{D}$ is an unrestricted matrix and then used the chain rule to obtain the derivatives over $\mathrm{vec}(\boldsymbol{L})$, deleting the entries that correspond to the entries in $\boldsymbol{L}$ that are 0 by definition. The vector of the first-order partial derivatives with respect to $\mathrm{vec}(\boldsymbol{D})$ are

$$\begin{aligned}\mathcal{U}_j\{\mathrm{vec}(\boldsymbol{D})\} &= -\frac{a_{0j}}{2}\frac{\partial\log|\boldsymbol{V}_{(j)}|}{\partial\,\mathrm{vec}(\boldsymbol{D})} - \frac{a_{0j}}{2\sigma^2}\left\{\frac{\partial\boldsymbol{r}_{(j)}^\top\boldsymbol{V}_{(j)}^{-1}\boldsymbol{r}_{(j)}}{\partial\,\mathrm{vec}(\boldsymbol{D})} + \frac{\partial\mathrm{tr}(\boldsymbol{V}_{(j)}^{-1}\boldsymbol{\Omega}_{0(j)})}{\partial\,\mathrm{vec}(\boldsymbol{D})}\right\} \\ &-\frac{a_{0j}}{2}\frac{\frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\,\mathrm{vec}(\boldsymbol{D})}}{Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})} - \frac{a_{0j}}{2Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}\left[\frac{\partial\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2}{\partial\,\mathrm{vec}(\boldsymbol{D})} + \frac{\partial\mathrm{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)})}{\partial\,\mathrm{vec}(\boldsymbol{D})}\right] \\ &+\frac{a_{0j}}{2Var^2(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}\left[\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2 + \mathrm{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)})\right]\frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\,\mathrm{vec}(\boldsymbol{D})}. \qquad (3.16)\end{aligned}$$

To exemplify the calculations, we derive $\frac{\partial\log|\boldsymbol{V}_{(j)}|}{\partial\,\mathrm{vec}(\boldsymbol{D})}$. By the formula $\frac{\partial}{\partial x_j}\log|\boldsymbol{F}| =$

$\text{tr}(\boldsymbol{F}^{-1}\frac{\partial \boldsymbol{F}}{\partial x_j})$ (Harville, 1997, page 305), and using basic properties of the trace operator we have that the derivative with respect to the ikth element of $\boldsymbol{D}$ is

$$\begin{aligned}\frac{\partial \log|\boldsymbol{V}_{(j)}|}{\partial M_{ik}} &= \text{tr}\left(\boldsymbol{V}_{(j)}^{-1}\frac{\partial \boldsymbol{V}_{(j)}}{\partial M_{ik}}\right) = \text{tr}\left\{\boldsymbol{V}_{(j)}^{-1}\frac{\partial}{\partial M_{ik}}\left(\boldsymbol{I}_{(j)} + \boldsymbol{Z}_{(j)}\boldsymbol{D}\boldsymbol{Z}_{(j)}^{\top}\right)\right\} \\ &= \text{tr}\left(\boldsymbol{V}_{(j)}^{-1}\boldsymbol{Z}_{(j)}\frac{\partial \boldsymbol{D}}{\partial M_{ik}}\boldsymbol{Z}_{(j)}^{\top}\right) = \text{tr}\left(\boldsymbol{V}_{(j)}^{-1}\boldsymbol{Z}_{(j)}\boldsymbol{u}_i\boldsymbol{u}_k^{\top}\boldsymbol{Z}_{(j)}^{\top}\right) = \boldsymbol{u}_k^{\top}\boldsymbol{Z}_{(j)}^{\top}\boldsymbol{V}_{(j)}^{-1}\boldsymbol{Z}_{(j)}\boldsymbol{u}_i \\ &= [\boldsymbol{Z}_{(j)}^{\top}\boldsymbol{V}_{(j)}^{-1}\boldsymbol{Z}_{(j)}]_{ki},\end{aligned}$$

where $\boldsymbol{u}_k^{\top}$ and $\boldsymbol{u}_i$ are the kth row and ith column of the identity matrix, respectively. Thus, $\frac{\partial \log|\boldsymbol{V}_{(j)}|}{\partial M_{ik}}$ equals the kith element of $\boldsymbol{Z}_{(j)}^{\top}\boldsymbol{V}_{(j)}^{-1}\boldsymbol{Z}_{(j)}$. Expressing the derivatives in terms of the matrix $\boldsymbol{D}$ (i.e. getting a matrix of derivatives) results in

$$\frac{\partial \log|\boldsymbol{V}_{(j)}|}{\partial \boldsymbol{D}} = \left(\boldsymbol{Z}_{(j)}^{\top}\boldsymbol{V}_{(j)}^{-1}\boldsymbol{Z}_{(j)}\right)^{\top} = \boldsymbol{Z}_{(j)}^{\top}\boldsymbol{V}_{(j)}^{-1}\boldsymbol{Z}_{(j)},$$

as $\boldsymbol{V}_{(j)}$ is symmetric by definition. Thus, using vector notation, we get that

$$\frac{\partial \log|\boldsymbol{V}_{(j)}|}{\partial \,\text{vec}(\boldsymbol{D})} = \text{vec}(\boldsymbol{Z}_{(j)}^{\top}\boldsymbol{V}_{(j)}^{-1}\boldsymbol{Z}_{(j)}).$$

Similarly, it can be shown that $\frac{\partial \boldsymbol{r}_{(j)}^{\top}\boldsymbol{V}_{(j)}^{-1}\boldsymbol{r}_{(j)}}{\partial \,\text{vec}(\boldsymbol{D})} = -\,\text{vec}(\boldsymbol{v}_{(j)}\boldsymbol{v}_{(j)}^{\top})$ and $\frac{\partial \text{tr}(\boldsymbol{V}_{(j)}^{-1}\boldsymbol{\Omega}_{0(j)})}{\partial \,\text{vec}(\boldsymbol{D})} = -\,\text{vec}(\boldsymbol{C}_{(j)})$, with $\boldsymbol{v}_{(j)} = \boldsymbol{Z}_{(j)}^{\top}\boldsymbol{V}_{(j)}^{-1}\boldsymbol{r}_{(j)}$ and $\boldsymbol{C}_{(j)} = \boldsymbol{Z}_{(j)}^{\top}\boldsymbol{V}_{(j)}^{-1}\boldsymbol{\Omega}_{0(j)}\boldsymbol{V}_{(j)}^{-1}\boldsymbol{Z}_{(j)}$. To see this, note that we used the formula $\frac{\partial \boldsymbol{F}^{-1}}{\partial x_j} = -\boldsymbol{F}^{-1}\frac{\partial \boldsymbol{F}}{\partial x_j}\boldsymbol{F}^{-1}$ (Harville, 1997, page 307). Thus,

$$\begin{aligned}\frac{\partial}{\partial M_{ik}}\boldsymbol{r}_{(j)}^{\top}\boldsymbol{V}_{(j)}^{-1}\boldsymbol{r}_{(j)} &= \text{tr}\left(\boldsymbol{r}_{(j)}\boldsymbol{r}_{(j)}^{\top}\frac{\partial \boldsymbol{V}_{(j)}^{-1}}{\partial M_{ik}}\right) \\ &= -\text{tr}\left(\boldsymbol{r}_{(j)}\boldsymbol{r}_{(j)}^{\top}\boldsymbol{V}_{(j)}^{-1}\frac{\partial \boldsymbol{V}_{(j)}}{\partial M_{ik}}\boldsymbol{V}_{(j)}^{-1}\right) \\ &= -\text{tr}\left(\boldsymbol{r}_{(j)}\boldsymbol{r}_{(j)}^{\top}\boldsymbol{V}_{(j)}^{-1}\boldsymbol{Z}_{(j)}\boldsymbol{u}_i\boldsymbol{u}_k^{\top}\boldsymbol{Z}_{(j)}^{\top}\boldsymbol{V}_{(j)}^{-1}\right) \\ &= -\boldsymbol{u}_k^{\top}\boldsymbol{Z}_{(j)}^{\top}\boldsymbol{V}_{(j)}^{-1}\boldsymbol{r}_{(j)}\boldsymbol{r}_{(j)}^{\top}\boldsymbol{V}_{(j)}^{-1}\boldsymbol{Z}_{(j)}\boldsymbol{u}_i\end{aligned}$$

Thus, $\frac{\partial \boldsymbol{r}_{(j)}^{\top}\boldsymbol{V}_{(j)}^{-1}\boldsymbol{r}_{(j)}}{\partial \boldsymbol{D}} = -(\boldsymbol{v}_{(j)}\boldsymbol{v}_{(j)}^{\top})^{\top} = -(\boldsymbol{v}_{(j)}\boldsymbol{v}_{(j)}^{\top})$, which leads to $\frac{\partial \boldsymbol{r}_{(j)}^{\top}\boldsymbol{V}_{(j)}^{-1}\boldsymbol{r}_{(j)}}{\partial \,\text{vec}(\boldsymbol{D})} = -\,\text{vec}(\boldsymbol{v}_{(j)}\boldsymbol{v}_{(j)}^{\top})$. Next, for the derivatives in the time-to-dropout model, it turned out to be simpler to first calculate the derivatives with respect to $\text{vec}(\boldsymbol{B})$ ($\boldsymbol{B} = \boldsymbol{D}^{-1}$) and then get the derivatives with respect to $\text{vec}(\boldsymbol{D})$ by the chain rule. Note that, by the matrix formula $(\boldsymbol{R} + \boldsymbol{STU})^{-1} = \boldsymbol{R}^{-1} - \boldsymbol{R}^{-1}\boldsymbol{S}(\boldsymbol{T}^{-1} + \boldsymbol{UR}^{-1}\boldsymbol{S})^{-1}\boldsymbol{UR}^{-1}$, we have that

$\boldsymbol{\Gamma}_{(j)}^{-1} = (\boldsymbol{D}^{-1} + \boldsymbol{Z}_{(j)}^{\top}\boldsymbol{Z}_{(j)})^{-1} = \boldsymbol{D} - \boldsymbol{D}\boldsymbol{Z}_{(j)}^{\top}\boldsymbol{V}_{(j)}^{-1}\boldsymbol{Z}_{(j)}\boldsymbol{D}$, thus $Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})$ equals $\sigma^2(\nu^2 + \boldsymbol{\lambda}^{\top}\boldsymbol{\Gamma}_{(j)}^{-1}\boldsymbol{\lambda})$. Then in a very similar way, it can be shown that

$$\frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial \operatorname{vec}(\boldsymbol{B})} = -\sigma^2 \operatorname{vec}(\boldsymbol{\Gamma}_{(j)}^{-1}\boldsymbol{\lambda}\boldsymbol{\lambda}^{\top}\boldsymbol{\Gamma}_{(j)}^{-1}).$$

Next, by the formula $(\boldsymbol{P}^{-1} + \boldsymbol{B}^{\top}\boldsymbol{R}^{-1}\boldsymbol{B})^{-1}\boldsymbol{B}^{\top}\boldsymbol{R}^{-1} = \boldsymbol{P}\boldsymbol{B}^{\top}(\boldsymbol{B}\boldsymbol{P}\boldsymbol{B}^{\top} + \boldsymbol{R})^{-1}$, with $\boldsymbol{P}$ and $\boldsymbol{R}$ being positively definite matrices, it follows that $(\boldsymbol{D}^{-1} + \boldsymbol{Z}_{(j)}^{\top}\boldsymbol{Z}_{(j)})^{-1}\boldsymbol{Z}_{(j)}^{\top} = \boldsymbol{D}\boldsymbol{Z}_{(j)}^{\top}\boldsymbol{V}_{(j)}^{-1}$, which means that $E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta}) = \zeta + \boldsymbol{\lambda}^{\top}\boldsymbol{\Gamma}_{(j)}^{-1}\boldsymbol{Z}_{(j)}^{\top}\boldsymbol{r}_{(j)}$. Thus, by similar manipulations, we get that

$$\frac{\partial E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial \operatorname{vec}(\boldsymbol{B})} = -\operatorname{vec}(\boldsymbol{\Gamma}_{(j)}^{-1}\boldsymbol{\lambda}\boldsymbol{r}_{(j)}^{\top}\boldsymbol{Z}_{(j)}\boldsymbol{\Gamma}_{(j)}^{-1}),$$

while $\frac{\partial\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2}{\partial \operatorname{vec}(\boldsymbol{B})} = -2\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}\frac{\partial E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial \operatorname{vec}(\boldsymbol{B})}$. Also, it can be shown that $\frac{\partial \operatorname{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)})}{\partial \boldsymbol{B}} = -\boldsymbol{\Gamma}_{(j)}^{-1}\boldsymbol{\lambda}\boldsymbol{\lambda}^{\top}\boldsymbol{\Gamma}_{(j)}^{-1}\boldsymbol{Z}_{(j)}^{\top}\boldsymbol{\Omega}_{0(j)}\boldsymbol{Z}_{(j)}\boldsymbol{\Gamma}_{(j)}^{-1} - (\boldsymbol{\Gamma}_{(j)}^{-1}\boldsymbol{\lambda}\boldsymbol{\lambda}^{\top}\boldsymbol{\Gamma}_{(j)}^{-1}\boldsymbol{Z}_{(j)}^{\top}\boldsymbol{\Omega}_{0(j)}\boldsymbol{Z}_{(j)}\boldsymbol{\Gamma}_{(j)}^{-1})^{\top}$, and by the definition of the commutation matrix $\boldsymbol{K}_{mn}$, $\operatorname{vec}(\boldsymbol{A}^{\top}) = \boldsymbol{K}_{mn}\operatorname{vec}(\boldsymbol{A})$ with $\boldsymbol{A}$ being a $m \times n$ matrix (Harville, 1997, page 344), it follows that

$$\frac{\partial \operatorname{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)})}{\partial \operatorname{vec}(\boldsymbol{B})} = -(\boldsymbol{I}_{q^2} + \boldsymbol{K}_{qq})\operatorname{vec}(\boldsymbol{\Gamma}_{(j)}^{-1}\boldsymbol{\lambda}\boldsymbol{\lambda}^{\top}\boldsymbol{\Gamma}_{(j)}^{-1}\boldsymbol{Z}_{(j)}^{\top}\boldsymbol{\Omega}_{0(j)}\boldsymbol{Z}_{(j)}\boldsymbol{\Gamma}_{(j)}^{-1}).$$

Finally, for the time-to-dropout model, we need to also represent the derivatives in terms of $\operatorname{vec}(\boldsymbol{D})$, which can be carried out through the chain rule. Thus, if $f(\boldsymbol{B})$ is a scalar- or vector-valued function of $\boldsymbol{B}$, then

$$\begin{aligned}\frac{\partial f(\boldsymbol{B})}{\partial \operatorname{vec}(\boldsymbol{D})^{\top}} &= \frac{\partial f(\boldsymbol{B})}{\partial \operatorname{vec}(\boldsymbol{B})^{\top}}\frac{\partial \operatorname{vec}(\boldsymbol{B})}{\partial \operatorname{vec}(\boldsymbol{D})^{\top}} = \frac{\partial f(\boldsymbol{B})}{\partial \operatorname{vec}(\boldsymbol{B})^{\top}}\frac{\partial \operatorname{vec}(\boldsymbol{D}^{-1})}{\partial \operatorname{vec}(\boldsymbol{D})^{\top}} \\ &= -\frac{\partial f(\boldsymbol{B})}{\partial \operatorname{vec}(\boldsymbol{B})^{\top}}\left\{(\boldsymbol{D}^{-1})^{\top} \otimes \boldsymbol{D}^{-1}\right\}\frac{\partial \operatorname{vec}(\boldsymbol{D})}{\partial \operatorname{vec}(\boldsymbol{D})^{\top}} = -\frac{\partial f(\boldsymbol{B})}{\partial \operatorname{vec}(\boldsymbol{B})^{\top}}\left(\boldsymbol{D}^{-1} \otimes \boldsymbol{D}^{-1}\right),\end{aligned}$$

where $\otimes$ stands for the kronecker product of matrices. Note that if $f(\boldsymbol{B})$ is a scalar-valued function of $\boldsymbol{B}$, the derivative in terms of the column vector $\operatorname{vec}(\boldsymbol{D})$ is $\frac{\partial f(\boldsymbol{B})}{\partial \operatorname{vec}(\boldsymbol{D})} = -\left(\boldsymbol{D}^{-1} \otimes \boldsymbol{D}^{-1}\right)\frac{\partial f(\boldsymbol{B})}{\partial \operatorname{vec}(\boldsymbol{B})}$, as $\boldsymbol{D}$ is symmetric. Thus, denoting by $\phi(\boldsymbol{B})$ the matrix of the first-order partial derivatives over $\boldsymbol{B}$, we have that

$$\frac{\partial f(\boldsymbol{B})}{\partial \operatorname{vec}(\boldsymbol{D})} = -\left(\boldsymbol{D}^{-1} \otimes \boldsymbol{D}^{-1}\right)\operatorname{vec}\{\phi(\boldsymbol{B})\} = -\operatorname{vec}\{\boldsymbol{D}^{-1}\phi(\boldsymbol{B})\boldsymbol{D}^{-1}\},$$

by the formula $\operatorname{vec}(\boldsymbol{A}\boldsymbol{B}\boldsymbol{C}) = (\boldsymbol{C}^{\top} \otimes \boldsymbol{A})\operatorname{vec}(\boldsymbol{A})$ (Harville, 1997, page 341).

Moving on the calculation of the vec($\boldsymbol{D}$) block of the Hessian matrix, letting

$$f_j\{\text{vec}(\boldsymbol{D})\} = Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\left[\frac{\partial\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2}{\partial\,\text{vec}(\boldsymbol{D})} + \frac{\partial\text{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)})}{\partial\,\text{vec}(\boldsymbol{D})}\right] - \left[\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2 + \text{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)})\right]\frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\,\text{vec}(\boldsymbol{D})},$$

and $g_j\{\text{vec}(\boldsymbol{D})\} = Var^2(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})$, we have that the second-order partial derivatives over vec($\boldsymbol{D}$) are

$$\begin{aligned}\mathcal{H}_j\{\text{vec}(\boldsymbol{D})\} = &-\frac{a_{0j}}{2}\frac{\partial^2\log|\boldsymbol{V}_{(j)}|}{\partial\,\text{vec}(\boldsymbol{D})\partial\,\text{vec}(\boldsymbol{D})^\top} - \frac{a_{0j}}{2\sigma^2}\left\{\frac{\partial^2\boldsymbol{r}_{(j)}^\top\boldsymbol{V}_{(j)}^{-1}\boldsymbol{r}_{(j)}}{\partial\,\text{vec}(\boldsymbol{D})\partial\,\text{vec}(\boldsymbol{D})^\top} + \frac{\partial^2\text{tr}(\boldsymbol{V}_{(j)}^{-1}\boldsymbol{\Omega}_{0(j)})}{\partial\,\text{vec}(\boldsymbol{D})\partial\,\text{vec}(\boldsymbol{D})^\top}\right\}\\
&- \frac{a_{0j}}{2Var^2(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}\left\{Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\frac{\partial^2 Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\,\text{vec}(\boldsymbol{D})\,\text{vec}(\boldsymbol{D})^\top} - \frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\,\text{vec}(\boldsymbol{D})}\frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\,\text{vec}(\boldsymbol{D})^\top}\right\}\\
&- \frac{a_{0j}}{2}\left[g_j\{\text{vec}(\boldsymbol{D})\}\frac{\partial f_j\{\text{vec}(\boldsymbol{D})\}}{\partial\,\text{vec}(\boldsymbol{D})^\top} - f_j\{\text{vec}(\boldsymbol{D})\}\frac{\partial g_j\{\text{vec}(\boldsymbol{D})\}}{\partial\,\text{vec}(\boldsymbol{D})^\top}\right]\Big/ g_j^2\{\text{vec}(\boldsymbol{D})\},\end{aligned} \quad (3.17)$$

where $\frac{\partial g_j\{\text{vec}(\boldsymbol{D})\}}{\partial\,\text{vec}(\boldsymbol{D})^\top} = 2Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\,\text{vec}(\boldsymbol{D})^\top}$ and

$$\begin{aligned}\frac{\partial f_j\{\text{vec}(\boldsymbol{D})\}}{\partial\,\text{vec}(\boldsymbol{D})^\top} &= \left[\frac{\partial\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2}{\partial\,\text{vec}(\boldsymbol{D})} + \frac{\partial\text{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)})}{\partial\,\text{vec}(\boldsymbol{D})}\right]\frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\,\text{vec}(\boldsymbol{D})^\top}\\
&+ Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\left[\frac{\partial^2\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2}{\partial\,\text{vec}(\boldsymbol{D})\partial\,\text{vec}(\boldsymbol{D})^\top} + \frac{\partial^2\text{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)})}{\partial\,\text{vec}(\boldsymbol{D})\partial\,\text{vec}(\boldsymbol{D})^\top}\right]\\
&- \frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\,\text{vec}(\boldsymbol{D})}\left[\frac{\partial\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2}{\partial\,\text{vec}(\boldsymbol{D})^\top} + \frac{\partial\text{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)})}{\partial\,\text{vec}(\boldsymbol{D})^\top}\right]\\
&- \left[\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2 + \text{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)})\right]\frac{\partial^2 Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\,\text{vec}(\boldsymbol{D})\partial\,\text{vec}(\boldsymbol{D})^\top},\end{aligned}$$

where, through the product rule of differentiation,

$$\frac{\partial^2\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2}{\partial\,\text{vec}(\boldsymbol{D})\partial\,\text{vec}(\boldsymbol{D})^\top} = 2\left[\frac{\partial E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\,\text{vec}(\boldsymbol{D})}\frac{\partial E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\,\text{vec}(\boldsymbol{D})^\top} - \{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}\frac{\partial^2 E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\,\text{vec}(\boldsymbol{D})\partial\,\text{vec}(\boldsymbol{D})^\top}\right]. \quad (3.18)$$

To show the manipulations that need to be made in the above expressions, similarly to Lindstrom and Bates (1988), we derive $\frac{\partial^2\boldsymbol{r}_{(j)}^\top\boldsymbol{V}_{(j)}^{-1}\boldsymbol{r}_{(j)}}{\partial\,\text{vec}(\boldsymbol{D})\partial\,\text{vec}(\boldsymbol{D})^\top}$ using the product rule along

with standard properties of the vec operator and the kronecker product.

$$\begin{aligned}
\frac{\partial^2 \boldsymbol{r}_{(j)}^\top \boldsymbol{V}_{(j)}^{-1}\boldsymbol{r}_{(j)}}{\partial \operatorname{vec}(\boldsymbol{D})\partial \operatorname{vec}(\boldsymbol{D})^\top} &= -\frac{\partial \operatorname{vec}(\boldsymbol{v}_{(j)}\boldsymbol{v}_{(j)}^\top)}{\partial \operatorname{vec}(\boldsymbol{D})^\top} = -(\boldsymbol{v}_{(j)}\otimes \boldsymbol{I})\frac{\partial \operatorname{vec}(\boldsymbol{v}_{(j)})}{\partial \operatorname{vec}(\boldsymbol{D})^\top} - (\boldsymbol{I}\otimes \boldsymbol{v}_{(j)})\frac{\partial \operatorname{vec}(\boldsymbol{v}_{(j)}^\top)}{\partial \operatorname{vec}(\boldsymbol{D})^\top} \\
&= -(\boldsymbol{v}_{(j)}\otimes \boldsymbol{I})\frac{\partial \operatorname{vec}(\boldsymbol{Z}_{(j)}^\top \boldsymbol{V}_{(j)}^{-1}\boldsymbol{r}_{(j)})}{\partial \operatorname{vec}(\boldsymbol{D})^\top} - (\boldsymbol{I}\otimes \boldsymbol{v}_{(j)})\frac{\partial \operatorname{vec}(\boldsymbol{r}_{(j)}^\top \boldsymbol{V}_{(j)}^{-1}\boldsymbol{Z}_{(j)})}{\partial \operatorname{vec}(\boldsymbol{D})^\top} \\
&= -(\boldsymbol{v}_{(j)}\otimes \boldsymbol{I})(\boldsymbol{r}_{(j)}^\top \otimes \boldsymbol{Z}_{(j)}^\top)\frac{\partial \operatorname{vec}(\boldsymbol{V}_{(j)}^{-1})}{\partial \operatorname{vec}(\boldsymbol{D})^\top} \\
&- (\boldsymbol{I}\otimes \boldsymbol{v}_{(j)})(\boldsymbol{Z}_{(j)}^\top \otimes \boldsymbol{r}_{(j)}^\top)\frac{\partial \operatorname{vec}(\boldsymbol{V}_{(j)}^{-1})}{\partial \operatorname{vec}(\boldsymbol{D})^\top} \\
&= (\boldsymbol{v}_{(j)}\boldsymbol{r}_{(j)}^\top \otimes \boldsymbol{Z}_{(j)}^\top)(\boldsymbol{V}_{(j)}^{-1}\otimes \boldsymbol{V}_{(j)}^{-1})\frac{\partial \operatorname{vec}(\boldsymbol{Z}_{(j)}\boldsymbol{D}\boldsymbol{Z}_{(j)}^\top)}{\partial \operatorname{vec}(\boldsymbol{D})^\top} \\
&+ (\boldsymbol{Z}_{(j)}^\top \otimes \boldsymbol{v}_{(j)}\boldsymbol{r}_{(j)}^\top)(\boldsymbol{V}_{(j)}^{-1}\otimes \boldsymbol{V}_{(j)}^{-1})\frac{\partial \operatorname{vec}(\boldsymbol{Z}_{(j)}\boldsymbol{D}\boldsymbol{Z}_{(j)}^\top)}{\partial \operatorname{vec}(\boldsymbol{D})^\top} \\
&= (\boldsymbol{v}_{(j)}\boldsymbol{v}_{(j)}^\top \otimes \boldsymbol{Z}_{(j)}^\top \boldsymbol{V}_{(j)}^{-1}\boldsymbol{Z}_{(j)}) + (\boldsymbol{Z}_{(j)}^\top \boldsymbol{V}_{(j)}^{-1}\boldsymbol{Z}_{(j)}\otimes \boldsymbol{v}_{(j)}\boldsymbol{v}_{(j)}^\top).
\end{aligned}$$

Using similar arguments, it can be shown that $\frac{\partial^2 \log|\boldsymbol{V}_{(j)}|}{\partial \operatorname{vec}(\boldsymbol{D})\partial \operatorname{vec}(\boldsymbol{D})^\top} = -(\boldsymbol{Z}_{(j)}^\top \boldsymbol{V}_{(j)}^{-1}\boldsymbol{Z}_{(j)}\otimes \boldsymbol{Z}_{(j)}^\top \boldsymbol{V}_{(j)}^{-1}\boldsymbol{Z}_{(j)})$ and $\frac{\partial^2 \operatorname{tr}(\boldsymbol{V}_{(j)}^{-1}\boldsymbol{\Omega}_{0(j)})}{\partial \operatorname{vec}(\boldsymbol{D})\partial \operatorname{vec}(\boldsymbol{D})^\top} = (\boldsymbol{C}_{(j)}\otimes \boldsymbol{Z}_{(j)}^\top \boldsymbol{V}_{(j)}^{-1}\boldsymbol{Z}_{(j)}) + (\boldsymbol{Z}_{(j)}^\top \boldsymbol{V}_{(j)}^{-1}\boldsymbol{Z}_{(j)}\otimes \boldsymbol{C}_{(j)})$. As when calculating the first-order partial derivatives, the second derivative matrix of the terms associated with the time-to-dropout model can be easier expressed in terms of $\operatorname{vec}(\boldsymbol{B})$. In this case, similarly to the above manipulations, it can be shown that $\frac{\partial^2 Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial \operatorname{vec}(\boldsymbol{B})\partial \operatorname{vec}(\boldsymbol{B})^\top} = \sigma^2(\boldsymbol{\Gamma}_{(j)}^{-1}\boldsymbol{\lambda}\boldsymbol{\lambda}^\top \boldsymbol{\Gamma}_{(j)}^{-1}\otimes \boldsymbol{\Gamma}_{(j)}^{-1}) + \sigma^2(\boldsymbol{\Gamma}_{(j)}^{-1}\otimes \boldsymbol{\Gamma}_{(j)}^{-1}\boldsymbol{\lambda}\boldsymbol{\lambda}^\top \boldsymbol{\Gamma}_{(j)}^{-1})$,

$$\begin{aligned}
\frac{\partial^2 \operatorname{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)})}{\partial \operatorname{vec}(\boldsymbol{B})\partial \operatorname{vec}(\boldsymbol{B})^\top} &= (\boldsymbol{I}_{q^2}+\boldsymbol{K}_{qq})(\boldsymbol{\Gamma}_{(j)}^{-1}\boldsymbol{Z}_{(j)}^\top \boldsymbol{\Omega}_{0(j)}\boldsymbol{Z}_{(j)}\boldsymbol{\Gamma}_{(j)}^{-1}\boldsymbol{\lambda}\boldsymbol{\lambda}^\top \boldsymbol{\Gamma}_{(j)}^{-1}\otimes \boldsymbol{\Gamma}_{(j)}^{-1}) \\
&\quad (\boldsymbol{I}_{q^2}+\boldsymbol{K}_{qq})(\boldsymbol{\Gamma}_{(j)}^{-1}\boldsymbol{Z}_{(j)}^\top \boldsymbol{\Omega}_{0(j)}\boldsymbol{Z}_{(j)}\boldsymbol{\Gamma}_{(j)}^{-1}\otimes \boldsymbol{\Gamma}_{(j)}^{-1}\boldsymbol{\lambda}\boldsymbol{\lambda}^\top \boldsymbol{\Gamma}_{(j)}^{-1}) \\
&\quad (\boldsymbol{I}_{q^2}+\boldsymbol{K}_{qq})(\boldsymbol{\Gamma}_{(j)}^{-1}\otimes \boldsymbol{\Gamma}_{(j)}^{-1}\boldsymbol{\lambda}\boldsymbol{\lambda}^\top \boldsymbol{\Gamma}_{(j)}^{-1}\boldsymbol{Z}_{(j)}^\top \boldsymbol{\Omega}_{0(j)}\boldsymbol{Z}_{(j)}\boldsymbol{\Gamma}_{(j)}^{-1}),
\end{aligned}$$

and $\frac{\partial^2 E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial \operatorname{vec}(\boldsymbol{B})\partial \operatorname{vec}(\boldsymbol{B})^\top} = (\boldsymbol{\Gamma}_{(j)}^{-1}\boldsymbol{Z}_{(j)}^\top \boldsymbol{r}_{(j)}\boldsymbol{\lambda}^\top \boldsymbol{\Gamma}_{(j)}^{-1}\otimes \boldsymbol{\Gamma}_{(j)}^{-1}) + (\boldsymbol{\Gamma}_{(j)}^{-1}\otimes \boldsymbol{\Gamma}_{(j)}^{-1}\boldsymbol{\lambda}\boldsymbol{r}_{(j)}^\top \boldsymbol{Z}_{(j)}\boldsymbol{\Gamma}_{(j)}^{-1})$. Then, we need to re-express the second derivative matrix in terms of $\boldsymbol{D}$. Denoting by $f(\boldsymbol{B})$ the scalar function to be differentiated, since $\frac{\partial f(\boldsymbol{B})}{\partial \operatorname{vec}(\boldsymbol{D})^\top} = -\operatorname{vec}\{\boldsymbol{D}^{-1}\phi(\boldsymbol{B})\boldsymbol{D}^{-1}\}$, it easily follows that

$$\begin{aligned}
\frac{\partial^2 f(\boldsymbol{B})}{\partial \operatorname{vec}(\boldsymbol{D})\partial \operatorname{vec}(\boldsymbol{D})^\top} &= (\boldsymbol{D}^{-1}\otimes \boldsymbol{D}^{-1})\frac{\partial^2 f(\boldsymbol{B})}{\partial \operatorname{vec}(\boldsymbol{B})\partial \operatorname{vec}(\boldsymbol{B})^\top}(\boldsymbol{D}^{-1}\otimes \boldsymbol{D}^{-1}) \\
&+ \{\boldsymbol{D}^{-1}\phi(\boldsymbol{B})^\top \boldsymbol{D}^{-1}\otimes \boldsymbol{D}^{-1}\} + \{\boldsymbol{D}^{-1}\otimes \boldsymbol{D}^{-1}\phi(\boldsymbol{B})\boldsymbol{D}^{-1}\}.
\end{aligned}$$

Next, we give the $\{\boldsymbol{\beta}, \text{vec}(\boldsymbol{D})\}$ block of the Hessian matrix

$$\begin{aligned}\mathcal{H}\{\boldsymbol{\beta}, \text{vec}(\boldsymbol{D})\} &= \frac{a_{0j}}{2}\frac{\partial \boldsymbol{X}_{(j)}^{\top}\boldsymbol{V}_{(j)}^{-1}\boldsymbol{r}_{(j)}}{\partial \,\text{vec}(\boldsymbol{D})^{\top}} - \frac{a_{0j}}{2}\Bigg[Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\frac{\partial^2\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2}{\partial\boldsymbol{\beta}\partial\,\text{vec}(\boldsymbol{D})^{\top}} \\ &- \frac{\partial\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2}{\partial\boldsymbol{\beta}}\frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\,\text{vec}(\boldsymbol{D})^{\top}}\Bigg]/Var^2(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta}), \quad (3.19)\end{aligned}$$

where

$$\begin{aligned}\frac{\partial^2\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2}{\partial\boldsymbol{\beta}\partial\,\text{vec}(\boldsymbol{D})^{\top}} &= -2\frac{\partial\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}}{\partial\boldsymbol{\beta}}\frac{\partial E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\,\text{vec}(\boldsymbol{D})^{\top}} \\ &+2\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}\frac{\partial^2\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}}{\partial\boldsymbol{\beta}\partial\,\text{vec}(\boldsymbol{D})^{\top}}.\end{aligned}$$

For the terms associated with the time-to-dropout model, we again express the derivatives in terms of $\text{vec}(\boldsymbol{B})$ and then back-transform to the derivatives over $\text{vec}(\boldsymbol{D})$ via the chain rule. Using similar arguments, it can be shown that $\frac{\partial^2\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}}{\partial\boldsymbol{\beta}\partial\,\text{vec}(\boldsymbol{B})^{\top}} = -(\boldsymbol{\lambda}^{\top}\boldsymbol{\Gamma}_{(j)}^{-1}\otimes\boldsymbol{X}_{(j)}^{\top}\boldsymbol{Z}_{(j)}\boldsymbol{\Gamma}_{(j)}^{-1})$ and $\frac{\partial \boldsymbol{X}_{(j)}^{\top}\boldsymbol{V}_{(j)}^{-1}\boldsymbol{r}_{(j)}}{\partial\,\text{vec}(\boldsymbol{D})^{\top}} = -(\boldsymbol{v}_{(j)}^{\top}\otimes\boldsymbol{X}_{(j)}^{\top}\boldsymbol{V}_{(j)}^{-1}\boldsymbol{Z}_{(j)})$, which completes the calculation of the $\{\boldsymbol{\beta}, \text{vec}(\boldsymbol{D})\}$ block of the Hessian matrix.

Next, by similar manipulations, the $\{\text{vec}(\boldsymbol{D}), \zeta\}$ block of the Hessian matrix is

$$\begin{aligned}\mathcal{H}\{\text{vec}(\boldsymbol{D}), \zeta\} = \frac{a_{0j}}{Var^2(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}\Bigg[&- Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\frac{\partial E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\,\text{vec}(\boldsymbol{D})} \\ &-\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}\frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\,\text{vec}(\boldsymbol{D})}\Bigg]. \quad (3.20)\end{aligned}$$

This way, the $\{\text{vec}(\boldsymbol{D}), \phi\}$ block of the Hessian matrix is

$$\begin{aligned}\mathcal{H}\{\text{vec}(\boldsymbol{D}), \phi\} &= \frac{a_{0j}}{2}\frac{\frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\phi}\frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\,\text{vec}(\boldsymbol{D})}}{Var^2(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})} \\ &+ \frac{a_{0j}}{2}\frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\phi}Var^2(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\left[\frac{\partial\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2}{\partial\,\text{vec}(\boldsymbol{D})} + \frac{\partial\text{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)})}{\partial\,\text{vec}(\boldsymbol{D})}\right] \\ &- \frac{a_{0j}}{2}\frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{Var^4(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\partial\phi}\left[\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2 + \text{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)})\right]\times\frac{\partial Var^2(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\,\text{vec}(\boldsymbol{D})}. \quad (3.21)\end{aligned}$$

Carrying on with the calculation of $\{\boldsymbol{\lambda}, \text{vec}(\boldsymbol{D})\}$ block of the Hessian matrix, letting

$$\begin{aligned} f_j\{\text{vec}(\boldsymbol{D})\} &= Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta}) \left\{ \frac{\partial (u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2}{\partial \boldsymbol{\lambda}} + \frac{\partial \text{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)})}{\partial \boldsymbol{\lambda}} \right\} \\ &- \left[\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2 + \text{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)}) \right] \frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial \boldsymbol{\lambda}}, \end{aligned}$$

and $g_j\{\text{vec}(\boldsymbol{D})\} = Var^2(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})$, viewed as a function $\text{vec}(\boldsymbol{D})$, it can be shown that

$$\begin{aligned} \mathcal{H}\{\boldsymbol{\lambda}, \text{vec}(\boldsymbol{D})\} = -\frac{a_{0j}}{2} \frac{Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta}) \frac{\partial^2 Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial \boldsymbol{\lambda} \partial \,\text{vec}(\boldsymbol{D})^\top} - \frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial \boldsymbol{\lambda}} \frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial \,\text{vec}(\boldsymbol{D})^\top}}{Var^2(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})} \\ - \frac{a_{0j}}{2} \left[g_j\{\text{vec}(\boldsymbol{D})\} \frac{\partial f_j\{\text{vec}(\boldsymbol{D})\}}{\partial \,\text{vec}(\boldsymbol{D})^\top} + f_j\{\text{vec}(\boldsymbol{D})\} \frac{\partial g_j\{\text{vec}(\boldsymbol{D})\}}{\partial \,\text{vec}(\boldsymbol{D})^\top} \right] \Big/ g_j^2\{\text{vec}(\boldsymbol{D})\}, \quad (3.22) \end{aligned}$$

where $\frac{\partial g_j\{\text{vec}(\boldsymbol{D})\}}{\partial \,\text{vec}(\boldsymbol{D})^\top} = 2Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta}) \frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial \,\text{vec}(\boldsymbol{D})^\top}$ and

$$\begin{aligned} \frac{\partial f_j\{\text{vec}(\boldsymbol{D})\}}{\partial \,\text{vec}(\boldsymbol{D})^\top} &= \left[\frac{\partial \{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2}{\partial \boldsymbol{\lambda}} + \frac{\partial \text{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)})}{\partial \boldsymbol{\lambda}} \right] \frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial \,\text{vec}(\boldsymbol{D})^\top} \\ &+ Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta}) \left[\frac{\partial^2 \{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2}{\partial \boldsymbol{\lambda} \partial \,\text{vec}(\boldsymbol{D})^\top} + \frac{\partial^2 \text{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)})}{\partial \boldsymbol{\lambda} \partial \,\text{vec}(\boldsymbol{D})^\top} \right] \\ &- \frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial \boldsymbol{\lambda}} \left[\frac{\partial \{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2}{\partial \,\text{vec}(\boldsymbol{D})^\top} + \frac{\partial \text{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)})}{\partial \,\text{vec}(\boldsymbol{D})^\top} \right] \\ &- \left[\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2 + \text{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)}) \right] \frac{\partial^2 Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial \boldsymbol{\lambda} \partial \,\text{vec}(\boldsymbol{D})^\top}. \end{aligned}$$

It is then straightforward to show that $\frac{\partial^2 Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial \boldsymbol{\lambda} \partial \,\text{vec}(\boldsymbol{B})^\top} = -2\sigma^2(\boldsymbol{\lambda}\boldsymbol{\Gamma}_{(j)}^{-1} \otimes \boldsymbol{\Gamma}_{(j)}^{-1})$, $\frac{\partial^2 \{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}}{\partial \boldsymbol{\lambda} \partial \,\text{vec}(\boldsymbol{B})^\top} = (\boldsymbol{r}_{(j)}^\top \boldsymbol{Z}_{(j)} \boldsymbol{\Gamma}_{(j)}^{-1} \otimes \boldsymbol{\Gamma}_{(j)}^{-1})$ and

$$\begin{aligned} \frac{\partial \{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2}{\partial \boldsymbol{\lambda} \partial \,\text{vec}(\boldsymbol{D})^\top} &= 2 \Bigg[- \frac{\partial \{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}}{\partial \boldsymbol{\lambda}} \frac{\partial E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial \,\text{vec}(\boldsymbol{D})^\top} \\ &+ \{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\} \frac{\partial^2 \{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}}{\partial \boldsymbol{\lambda} \partial \,\text{vec}(\boldsymbol{D})^\top} \Bigg], \end{aligned}$$

completing the calculations required for the $\{\boldsymbol{\lambda}, \text{vec}(\boldsymbol{D})\}$ block of the Hessian matrix.

Using the reparameterization $\theta = \log(\sigma^2)$, the first-order partial derivatives with

respect to θ are

$$\begin{aligned}\mathcal{U}_j(\theta) &= -\frac{(j+1)a_{0j}}{2} + \frac{a_{0j}}{2e^{\theta}}\left\{\boldsymbol{r}_{(j)}^{\top}\boldsymbol{V}_{(j)}^{-1}\boldsymbol{r}_{(j)} + \text{tr}(\boldsymbol{V}_{(j)}^{-1}\boldsymbol{\Omega}_{0(j)})\right\} \\ &+ \frac{a_{0j}\frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\theta}}{2Var^2(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}\left[\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2 + \text{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)})\right], \quad (3.23)\end{aligned}$$

and the second derivative matrix are

$$\begin{aligned}\mathcal{H}(\theta) &= -\frac{a_{0j}}{2e^{\theta}}\left\{\boldsymbol{r}_{(j)}^{\top}\boldsymbol{V}_{(j)}^{-1}\boldsymbol{r}_{(j)} + \text{tr}(\boldsymbol{V}_{(j)}^{-1}\boldsymbol{\Omega}_{0(j)})\right\} - \frac{a_{0j}}{2}\left[\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2 + \text{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)})\right] \\ &\times \left\{Var^2(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\frac{\partial^2 Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\theta\partial\theta}\right. \\ &\left. - \frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\theta}\frac{\partial Var^2(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\theta}\right\} \Big/ Var^4(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta}), \quad (3.24)\end{aligned}$$

with $\frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\theta} = Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})$.

The $(\boldsymbol{\lambda},\theta)$ block of the Hessian matrix is

$$\begin{aligned}\mathcal{H}(\boldsymbol{\lambda},\theta) &= +\frac{a_{0j}}{2}\frac{\frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\theta}}{Var^2(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}\left[\frac{\partial\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2}{\partial\boldsymbol{\lambda}} + \frac{\partial\text{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)})}{\partial\boldsymbol{\lambda}}\right] \\ &+ \frac{a_{0j}}{2}\left[\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2 + \text{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)})\right] \\ &\times \frac{Var^2(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\frac{\partial^2 Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\boldsymbol{\lambda}\partial\theta} - \frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\boldsymbol{\lambda}}\frac{\partial Var^2(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\theta}}{Var^4(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}, \quad (3.25)\end{aligned}$$

where it can be easily shown that $\frac{\partial^2 Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\boldsymbol{\lambda}\partial\theta} = \frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\boldsymbol{\lambda}}$.

Next, the $(\boldsymbol{\beta},\theta)$ block of the Hessian matrix is

$$\mathcal{H}(\boldsymbol{\beta},\theta) = \frac{a_{0j}\frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\theta}}{2Var^2(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}\frac{\partial\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2}{\partial\boldsymbol{\beta}} - \frac{a_{0j}}{e^{\theta}}\boldsymbol{X}_{(j)}^{\top}\boldsymbol{V}_{(j)}^{-1}\boldsymbol{r}_{(j)}. \quad (3.26)$$

The (ζ,θ) block of the Hessian matrix is

$$\mathcal{H}(\zeta,\theta) = -a_{0j}\frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\theta}\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}/Var^2(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta}). \quad (3.27)$$

Similarly, the (ϕ,θ) block of the Hessian matrix can be shown to be equal to

$$\begin{aligned}\mathcal{H}(\phi,\theta) &= \frac{a_{0j}}{2}\frac{Var^2(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\frac{\partial^2 Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\phi\partial\theta} - \frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\phi}\frac{\partial Var^2(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\theta}}{Var^4(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})} \\ &\times \left[\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2 + \text{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)})\right]. \quad (3.28)\end{aligned}$$

Finally, the $\{\text{vec}(\boldsymbol{D}), \theta\}$ block of the Hessian matrix is equal to

$$\begin{aligned}\mathcal{H}\{\text{vec}(\boldsymbol{D}), \theta\} = &-\frac{a_{0j}}{2\sigma^2}\left\{\frac{\partial \boldsymbol{r}_{(j)}^{\top}\boldsymbol{V}_{(j)}^{-1}\boldsymbol{r}_{(j)}}{\partial \,\text{vec}(\boldsymbol{D})} + \frac{\partial \text{tr}(\boldsymbol{V}_{(j)}^{-1}\boldsymbol{\Omega}_{0(j)})}{\partial \,\text{vec}(\boldsymbol{D})}\right\} \\ &+\frac{a_{0j}}{2}\frac{\frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\theta}}{Var^2(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}\left[\frac{\partial\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2}{\partial \,\text{vec}(\boldsymbol{D})} + \frac{\partial \text{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)})}{\partial \,\text{vec}(\boldsymbol{D})}\right] \\ &+\frac{a_{0j}}{2}\left[\{u_j - E(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\}^2 + \text{tr}(\boldsymbol{A}_{(j)}\boldsymbol{\Omega}_{0(j)})\right] \\ &\times\frac{Var^2(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})\frac{\partial^2 Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial \,\text{vec}(\boldsymbol{D})\partial\theta} - \frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial \,\text{vec}(\boldsymbol{D})}\frac{\partial Var^2(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial\theta}}{Var^4(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})},\end{aligned} \tag{3.29}$$

where $\frac{\partial^2 Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial \,\text{vec}(\boldsymbol{D})\partial\theta} = \frac{\partial Var(U|\boldsymbol{\mu}_{0(j)};\boldsymbol{\theta})}{\partial \,\text{vec}(\boldsymbol{D})}$. We now need to express all the derivatives involving $\text{vec}(\boldsymbol{D})$ in terms of $\text{vec}(\boldsymbol{L})$, the Choleski factor of $\text{vec}(\boldsymbol{D})$. Assuming that $f(\boldsymbol{D})$ is a scalar- or vector-valued function of $\boldsymbol{D}$, it follows through the chain rule that

$$\frac{\partial f(\boldsymbol{D})}{\partial \,\text{vec}(\boldsymbol{L})^{\top}} = \frac{\partial f(\boldsymbol{D})}{\partial \,\text{vec}(\boldsymbol{D})^{\top}}\frac{\partial \,\text{vec}(\boldsymbol{D})}{\partial \,\text{vec}(\boldsymbol{L})^{\top}}.$$

Using the product rule and the very definition of the commutation matrix, we have that $\frac{\partial \,\text{vec}(\boldsymbol{D})}{\partial \,\text{vec}(\boldsymbol{L})^{\top}} = (\boldsymbol{L}^{\top}\otimes\boldsymbol{I}_q)\boldsymbol{K}_{qq} + (\boldsymbol{I}_q\otimes\boldsymbol{L}^{\top})$. Then, assuming that $f(\boldsymbol{D})$ is scalar and denoting by $\phi(\boldsymbol{D})$ the matrix of the partial derivatives over $\boldsymbol{D}$, it can be shown that

$$u\{\text{vec}(\boldsymbol{L})\} = \frac{\partial f(\boldsymbol{D})}{\partial \,\text{vec}(\boldsymbol{L})} = \text{vec}\{\boldsymbol{L}\phi(\boldsymbol{D})^{\top}\} + \text{vec}\{\boldsymbol{L}\phi(\boldsymbol{D})\},$$

whereas, by similar manipulations the second derivative matrix is equal to

$$\begin{aligned}\frac{\partial u\{\text{vec}(\boldsymbol{L})\}}{\partial \,\text{vec}(\boldsymbol{L})^{\top}} &= (\boldsymbol{I}_q\otimes\boldsymbol{L})\left\{(K_{qq}+\boldsymbol{I}_{q^2})\frac{\partial^2 f(\boldsymbol{D})}{\partial \,\text{vec}(\boldsymbol{D})\partial \,\text{vec}(\boldsymbol{D})^{\top}}\right\}\{(\boldsymbol{L}^{\top}\otimes\boldsymbol{I}_q)\boldsymbol{K}_{qq} + (\boldsymbol{I}_q\otimes\boldsymbol{L}^{\top})\} \\ &+ \{\phi(\boldsymbol{D}) + \phi(\boldsymbol{D})^{\top}\}\otimes\boldsymbol{I}_q.\end{aligned}$$

We compared the results from our analytical derivative formulas with those obtained from numerical approximation methods implemented in *numDeriv* package in *R* (Gilbert and Varadhan, 2016), yielding negligible differences of order less than 10^{-7}.

In the following subsections we define the true mechanisms with respect to which bias is quantified. It should be emphasized that the terms a_{0j}, $\boldsymbol{\mu}_{0(j)}$, and $\boldsymbol{\Omega}_{0(j)}$ are considered constant in Equation (3.5) as differentiation is taken over $\boldsymbol{\theta}$.

3.2.2 MAR Drop-out Completely Determined by Observed Measurements

Let us assume that subjects drop out of the study when the marker drops below some threshold value c. This scenario is based on the WHO guidelines up to 2015 that cART should be initiated when CD4 counts fall below 500 cells/μL. Note that $M = j$ implies that $Y_1 \geq c, \ldots, Y_{j-1} \geq c, Y_j < c$ for $j < Q$ and $Y_1 \geq c, \ldots, Y_{Q-1} \geq c, Y_Q \in \mathrm{R}$ for $M = Q$. As the marginal distribution of $\boldsymbol{Y}_{(j)}$ is multivariate normal and $M = j$ poses the above constraints, the distribution of $\boldsymbol{Y}_{(j)}|M = j$ is a truncated multivariate normal (Wilhelm and Manjunath, 2015). To calculate its moments, we used the method proposed by Wilhelm and Manjunath (2015), whereas the true probability of the jth dropout pattern, a_{0j}, was calculated by the cumulative distribution function of the multivariate normal distribution. In this dropout mechanism, the probability of dropout is completely dependent on c since $\Pr(M = j|M \geq j, \boldsymbol{Y}_{(j)}; \boldsymbol{\theta}_{t0}) = I(Y_j < c)$ for $j < Q$, which implies that increasing the value of c leads to higher rate of dropout (i.e. dropout occurs sooner).

Finally, we solved Equation (3.5) using the Newton-Raphson algorithm, where the last term in Equation (3.5) was approximated by quasi-Monte Carlo integration, a derandomized alternative to Monte Carlo integration (e.g. Morokoff and Caflisch, 1995), with 10^5 contributing points. $\boldsymbol{D}$ was parameterized in terms of its Choleski factor $\boldsymbol{L}$ (i.e. $\boldsymbol{L}^\top \boldsymbol{L} = \boldsymbol{D}$), with derivatives over vec($\boldsymbol{L}$) obtained from the chain rule. A detailed derivation is given in subsection 3.2.1. The results at convergence of the algorithm constitute the values to which the LN-SREM(RE) estimators converge. By comparing these values with the true parameter values, we calculated the bias in the LN-SREM(RE) model.

3.2.3 MAR Drop-out Stochastically Determined by Observed Measurements

We now generalize the previous true dropout mechanism by allowing the hazard of dropout to depend on the current observed marker levels via a probabilistic model.

Specifically, we assume a logistic model of the form $P(M = j|M \geq j, \boldsymbol{Y}_{(j)}; \boldsymbol{\theta}_{t0}) = \frac{e^{c_1+c_2(Y_j-Y^\star)}}{1+e^{c_1+c_2(Y_j-Y^\star)}}$, for $j < Q$, and $P(M = Q|M \geq Q, \boldsymbol{Y}_{(Q)}; \boldsymbol{\theta}_{t0}) = 1$, which translates into the corresponding probabilities being equal to the products

$$\begin{aligned} P(M = j|\boldsymbol{Y}_{(j)}; \boldsymbol{\theta}_{t0}) &= \prod_{k=1}^{j} \left\{1 + e^{c_1+c_2(Y_k-Y^\star)}\right\}^{-1} e^{c_1+c_2(Y_j-Y^\star)} \\ P(M = Q|\boldsymbol{Y}_{(Q)}; \boldsymbol{\theta}_{t0}) &= \prod_{k=1}^{Q-1} \left\{1 + e^{c_1+c_2(Y_k-Y^\star)}\right\}^{-1}, \end{aligned}$$

respectively. The parameter c_1 is associated with the chance of dropout when the marker is at a certain level $Y^\star$, whereas c_2 quantifies the change in the log-odds of dropout associated with one unit increase in the marker values. Thus, a zero value of c_2 corresponds to an MCAR dropout mechanism. Instead, if c_2 is non-zero, the dropout mechanism is MAR. To calculate the moments of the true distribution of $\boldsymbol{Y}_{(j)}|M = j; \boldsymbol{\theta}_{t0}$ and the marginal probabilities of dropout, a_{0j}, we used quasi-Monte Carlo integration along with importance sampling, using a multivariate normal importance density with mean equal to the mode of $\boldsymbol{Y}_{(j)}|M = j; \boldsymbol{\theta}_{t0}$ and covariance matrix equal to the inverse curvature at the mode.

First, recall that a_{0j} is equal to

$$a_{0j} = \Pr(M = j; \boldsymbol{\theta}_0) = \int \Pr(M = j|\boldsymbol{Y}_{(j)}; \boldsymbol{\theta}_{t0}) f(\boldsymbol{Y}_{(j)}; \boldsymbol{\theta}_0) d\boldsymbol{Y}_{(j)}.$$

which cannot be computed analytically unless the dropout is completely at random. This, however, can be written equivalently as

$$\int \Pr(M = j|\boldsymbol{Y}_{(j)}; \boldsymbol{\theta}_{t0}) \frac{f(\boldsymbol{Y}_{(j)}; \boldsymbol{\theta}_0)}{g(\boldsymbol{Y}_{(j)})} g(\boldsymbol{Y}_{(j)}) d\boldsymbol{Y}_{(j)},$$

where $g(\boldsymbol{Y}_{(j)})$ density the density of a multivariate Normal distribution with mean equal to the mode of $\boldsymbol{Y}_{(j)}|M = j; \boldsymbol{\theta}_0$ and covariance matrix equal to the inverse curvature at the mode. Therefore, to estimate a_{0j}, we used the formula

$$\widehat{a}_j = \frac{1}{N_{mc}} \sum_{i=1}^{N_{mc}} \Pr(M = j|\boldsymbol{Y}^i_{(j)}; \boldsymbol{\theta}_{t0}) \frac{f(\boldsymbol{Y}^i_{(j)}; \boldsymbol{\theta}_0)}{g(\boldsymbol{Y}^i_{(j)})},$$

where $\boldsymbol{Y}^i_{(j)}$s were obtained by first transforming the sobol sequences to univariate normal low-discrepancy sequences (Christophe and Petr, 2019) and then applying a

Choleski-based transformation to get at the multivariate normal distribution of $g(\boldsymbol{Y}_{(j)})$. The advantage of this approach is that if the conditional distribution of $\boldsymbol{Y}_{(j)}|M=j;\boldsymbol{\theta}_0$ is well approximated by a normal distribution, the ratio $\Pr(M=j|\boldsymbol{Y}_{(j)};\boldsymbol{\theta}_{t0})\frac{f(\boldsymbol{Y}_{(j)};\boldsymbol{\theta}_0)}{g(\boldsymbol{Y}_{(j)})}$ will be nearly constant in $\boldsymbol{Y}_{(j)}$, resulting in much more precise estimates. To estimate the expected value of $\boldsymbol{Y}_{(j)}|M=j;\boldsymbol{\theta}_0$, $\boldsymbol{\mu}_{0(j)}=\int \boldsymbol{Y}_{(j)}f(\boldsymbol{Y}_{(j)}|M=j;\boldsymbol{\theta}_0)d\boldsymbol{Y}_{(j)}$, we used the formula

$$\frac{1}{N_{mc}}\sum_{i=1}^{N_{mc}}\boldsymbol{Y}^i_{(j)}\frac{\Pr(M=j|\boldsymbol{Y}^i_{(j)};\boldsymbol{\theta}_{t0})f(\boldsymbol{Y}^i_{(j)};\boldsymbol{\theta}_0)}{g(\boldsymbol{Y}^i_{(j)})\widehat{a}_j}.$$

A similar formula was used to approximate $E(\boldsymbol{Y}_{(j)}\boldsymbol{Y}^\top_{(j)}|M=j;\boldsymbol{\theta}_0)$. Then the covariance matrix $\boldsymbol{\Omega}_{0(j)}$ can be estimated using the identity $E(\boldsymbol{Y}_{(j)}\boldsymbol{Y}^\top_{(j)}|M=j;\boldsymbol{\theta}_0)-\boldsymbol{\mu}_{0(j)}\boldsymbol{\mu}^\top_{0(j)}$. Using $N_{mc}=5\cdot 10^5$ draws resulted in the desired accuracy.

3.2.4 Parameters of the Data Generating Process and Bias Results

Throughout this analysis, all features regarding the true mechanisms generating the data were based on the CD4 count evolution during the HIV natural history, using results from models applied to data from the CASCADE study. Specifically, we assumed a random intercept and slope model for the square root of CD4 counts, with population parameter $\boldsymbol{\beta}_0=(23.60,-1.30)$; the variances of the random intercepts and slopes being equal to 22.6 and 1.85, respectively, and the covariance being equal to -2.07. The within-subject variance was set to 5.3.

To quantify the bias in the LN-SREM(RE) model, we solved Equation (3.5) for various values of the parameters of the dropout mechanisms. In the case of MAR dropout completely determined by observed measurements, the threshold c, below which the next measurements are missing, ranged from 50 to 550 cells/μL. Results are shown on the left hand side at the top of Figure 3.1. It is clear that the bias in the population slope is an increasing function of c (range: 2-322%), implying that the higher the dropout probability, the larger the bias in the estimated slope.

For the stochastic MAR dropout mechanism, we set c_1 to be the hazard of dropout at 500 cells/μL (on the logit scale), which approximately corresponds to the CD4 counts at HIV seroconversion (i.e. the "baseline"). We calculated the asymptotic bias in the

estimated slope for various values of the parameters (c_1, c_2). The results are presented on the left hand side at the bottom of Figure 3.1. When $c_2 = 0$, which corresponds to an MCAR mechanism, the LN-SREM(RE) model is nearly unbiased. For a given baseline dropout rate, the bias increases as c_2 decreases (increases in absolute value), while for a given value of c_2, the bias increases as the baseline hazard of dropout increases. Thus, provided that the dropout mechanism is not MCAR, the bias in the estimated slope increases as the dropout becomes heavier.

3.3 Asymptotic Bias in the PH-SREM(CV) when Fitted to MAR Data

In this section we study the potential bias in the PH-SREM(CV) model in which the hazard of dropout, modeled by a proportional hazards model, depends on the "true" current marker value (Wulfsohn and Tsiatis, 1997). Marker's trends are again modeled through a random intercept and slope model, whereas the proportional hazards model for the dropout mechanism is of the form

$$h\{t|m(t);\boldsymbol{\theta}_t\} = h_0(t;\boldsymbol{\psi})\exp\{\alpha m(t)\},$$

where $m(t) = (\beta_0 + b_0) + (\beta_1 + b_1)t$ is the underlying "true" marker value at time t and $h_0(t;\boldsymbol{\psi})$ denotes the baseline hazard function. $\log\{h_0(t;\boldsymbol{\psi})\}$ was modeled through restricted cubic splines of log time (e.g. Crowther et al., 2012) with two internal knots. Briefly, restricted cubic splines are cubic splines constrained to be linear beyond some boundary knots k_{min} and k_{max}, with such knots usually placed at the extreme observed log time values. Additionally, m distinct internal knots $k_1 < \ldots < k_m$ are specified, where $k_1 > k_{min}$ and $k_m < k_{max}$. A restricted cubic spline may then be written as

$$s(x;\boldsymbol{\psi}) = \psi_0 + \psi_1 x + \psi_2 v_1(x) + \ldots \psi_{m+1} v_m(x) = \boldsymbol{V}^\top(x)\boldsymbol{\psi},$$

where $x = \log(t)$ and $\boldsymbol{\psi}$ is an associated parameter vector. The jth basis function is defined for $j = 1, 2, \ldots, m$ as

$$v_j(x) = (x - k_j)_+^3 - \lambda_j(x - k_{min})_+^3 - (1 - \lambda_j)(x - k_{max})_+^3, \tag{3.30}$$

3. PERFORMANCE OF LONGITUDINAL AND TIME-TO-DROPOUT JOINT MODELS WHEN THE DROPOUT MECHANISM IS AT RANDOM

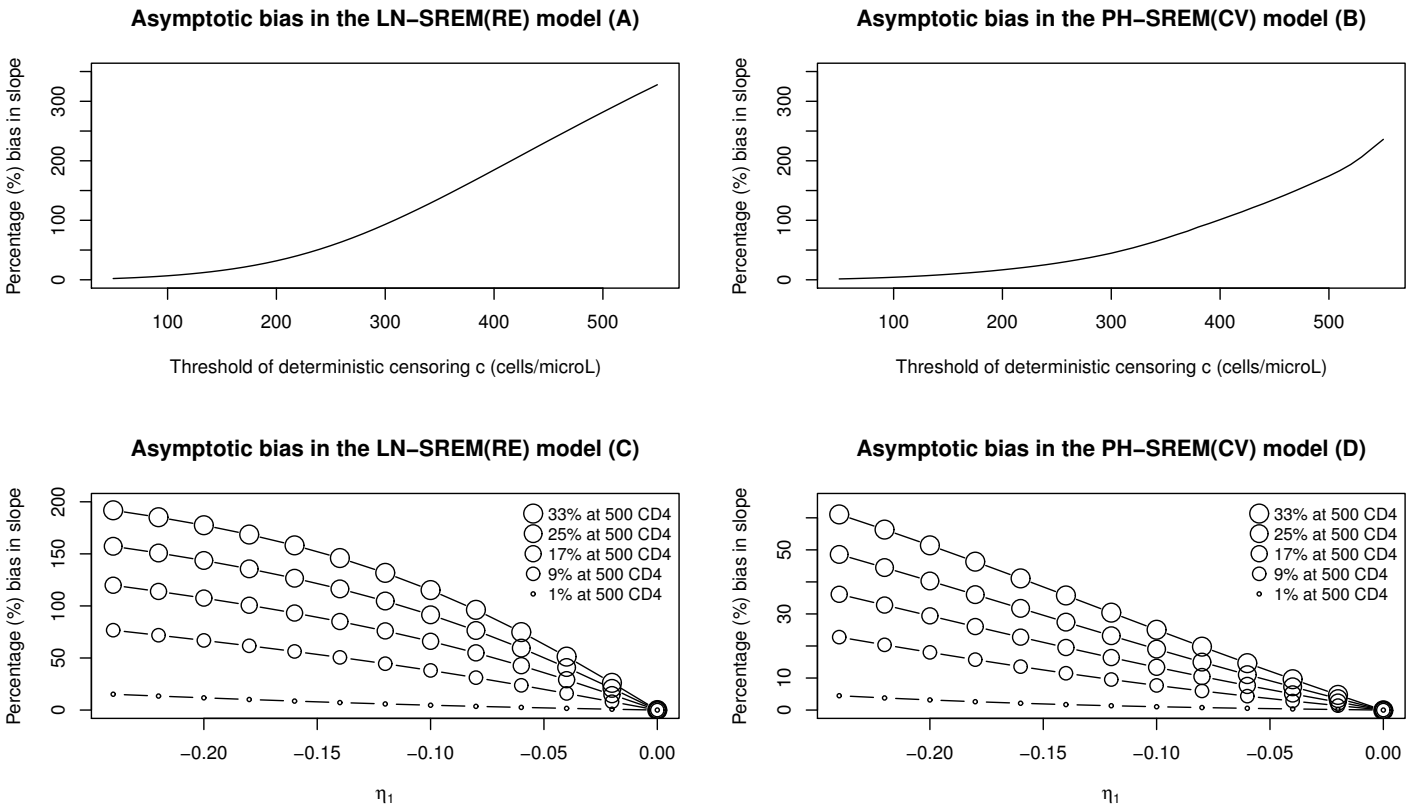


Figure 3.1: Asymptotic bias in a disease's marker rate of change estimated by the LN-SREM(RE) model (A and C) and the PH-SREM(CV) model (B and D) assuming MAR dropout mechanisms: subjects drop out when the marker reaches a certain threshold c (A and B; deterministic dropout) or when the probability of dropout is a function of marker values (C and D; stochastic dropout). In the latter case, c_2 measures the change in the log-odds of dropout associated with one unit decrease in the current marker value. The size of the plotting symbol is proportional to the dropout probability at 500 CD4 cells/μL, i.e. approximately the CD4 counts at seroconversion (baseline).

where $\lambda_j = (k_{max} - k_j)/(k_{max} - k_{min})$ and $(x)_+ = \max\{0, x\}$. The parameter α quantifies the dependence between the two submodels. Given $M = j$, the likelihood contribution is $\int f(\boldsymbol{Y}_{(j)}|\boldsymbol{b};\boldsymbol{\theta}_L)h(T|m(T);\boldsymbol{\theta}_t)^{I(j<Q)}S\{T|M_i(T);\boldsymbol{\theta}_t\}f(\boldsymbol{b};\boldsymbol{\theta}_L)d\boldsymbol{b}$, where $T = t_j + 0.1$ for $M < Q$ and $T = t_Q$ for $M = Q$, and $\boldsymbol{\theta}^\top = (\boldsymbol{\beta}^\top, \alpha, \text{vech}(\boldsymbol{D})^\top, \boldsymbol{\psi}, \sigma^2)$ is the parameter vector of the PH-SREM(CV) model. The conditional survival function, $S(t|\boldsymbol{b};\boldsymbol{\theta})$, equals

$$S\{t|M_i(t);\boldsymbol{\theta}_t\} = \exp\left(-\int_0^T \exp\left[s(\log(u);\boldsymbol{\psi}) + \alpha m(u)\right] du\right), \tag{3.31}$$

which is not available in closed form, though. To approximate this integral, we used the 15-point Gauss-Kronrod rule (Press et al., 2007).

To derive the asymptotic bias, similarly to Equation (3.5), we need to calculate expectations such as $E_{\boldsymbol{Y}_{(j)}|M=j;\boldsymbol{\theta}_0}\left\{\log f(\boldsymbol{Y}_{(j)}, T, \delta;\boldsymbol{\theta})|M = j;\boldsymbol{\theta}_0\right\}$, $j = 1, \ldots, Q$, with the expectation taken with respect to the true marker distribution given the dropout pattern. Since the likelihood cannot be expressed in closed-form, the expectation cannot be evaluated analytically. However, by making a simple but reasonable approximation to the likelihood, we can easily approximate the bias.

The likelihood of the PH-SREM(CV) model can be factorized as $f(\boldsymbol{Y}_{(j)}|T, \delta;\boldsymbol{\theta})f(T, \delta;\boldsymbol{\theta})$, where $f(T, \delta;\boldsymbol{\theta}) = \int f(T, \delta|\boldsymbol{b};\boldsymbol{\theta})f(\boldsymbol{b};\boldsymbol{\theta})d\boldsymbol{b}$ can be easily evaluated by adaptive Gauss-Hermite rules. The expected value and the covariance matrix of $\boldsymbol{Y}_{(j)}|T, \delta$ are equal to $\boldsymbol{\mu}_{(j)} = \boldsymbol{X}_{(j)}\boldsymbol{\beta} + \boldsymbol{Z}_{(j)}E(\boldsymbol{b}|T, \delta;\boldsymbol{\theta})$ and $\boldsymbol{\Omega}_{(j)} = \sigma^2\boldsymbol{I}_{(j)} + \boldsymbol{Z}_{(j)}Var(\boldsymbol{b}|T, \delta;\boldsymbol{\theta})\boldsymbol{Z}_{(j)}^\top$, respectively, which can be effectively evaluated using the same quadrature rules as those used in evaluating $f(T, \delta;\boldsymbol{\theta})$ (Saha and Jones, 2005). Making a Laplace-type approximation in which the distribution of $\boldsymbol{Y}_{(j)}|T, \delta$ is approximated by $N(\boldsymbol{\mu}_{(j)}, \boldsymbol{\Omega}_{(j)})$, the approximate log-likelihood contribution for the jth dropout pattern becomes a quadratic form in $\boldsymbol{Y}_{(j)}$. Upon observing that $f(T, \delta;\boldsymbol{\theta})$, as well as $\boldsymbol{\mu}_{(j)}$ and $\boldsymbol{\Omega}_{(j)}$, do not depend on $\boldsymbol{Y}_{(j)}$, the equation to be solved simplifies to

$$\frac{\partial}{\partial\boldsymbol{\theta}}\sum_{j=1}^{Q} a_{0j}\left\{-\frac{(\boldsymbol{\mu}_{0(j)} - \boldsymbol{\mu}_{(j)})^\top\boldsymbol{\Omega}_{(j)}^{-1}(\boldsymbol{\mu}_{0(j)} - \boldsymbol{\mu}_{(j)})}{2}\right.$$
$$\left.-\frac{\text{tr}(\boldsymbol{\Omega}_{(j)}^{-1}\boldsymbol{\Omega}_{0(j)})}{2} - \frac{\log|\boldsymbol{\Omega}_{(j)}|}{2} + \log f(T, \delta;\boldsymbol{\theta})\right\} = \boldsymbol{0}, \tag{3.32}$$

recalling that $\boldsymbol{\mu}_{0(j)}$ and $\boldsymbol{\Omega}_{0(j)}$ denote the mean and covariance matrix of the true distribution of $\boldsymbol{Y}_{(j)}|M = j;\boldsymbol{\theta}_0$, respectively, with $\boldsymbol{\mu}_{(j)}$ and $\boldsymbol{\Omega}_{(j)}$ referring to the corresponding model-based ones by the PH-SREM(CV) model. Nevertheless, as Equation (3.32) relies on the quality of the approximation made to the distribution of $\boldsymbol{Y}_{(j)}|T,\delta$, results obtained by solving Equation (3.32) may not be as accurate as desired. Thus, we proceeded with a better approach, using the results from Equation (3.32) only as starting values. Namely, we first factorized the likelihood as $f(\boldsymbol{Y}_{(j)};\boldsymbol{\theta})f(T,\delta|\boldsymbol{Y}_{(j)};\boldsymbol{\theta})$. Then the expectation of the log-likelihood of the marker model with respect to the true data distribution can be obtained analytically as $\log f(\boldsymbol{Y}_{(j)};\boldsymbol{\theta})$ is a quadratic form in $\boldsymbol{Y}_{(j)}$, whereas $f(T,\delta|\boldsymbol{Y}_{(j)};\boldsymbol{\theta}) = \int f(T,\delta|\boldsymbol{b};\boldsymbol{\theta})f(\boldsymbol{b}|\boldsymbol{Y}_{(j)};\boldsymbol{\theta})d\boldsymbol{b}$, with the distribution of $\boldsymbol{b}|\boldsymbol{Y}_{(j)}$ being multivariate normal with covariance matrix $\boldsymbol{\Sigma}_{b,(j)} = \sigma^2(\boldsymbol{D}^{-1}+\boldsymbol{Z}_{(j)}^\top\boldsymbol{Z}_{(j)})^{-1}$ and mean $\boldsymbol{\mu}_{b,(j)} = \boldsymbol{\Sigma}_{b,(j)}\boldsymbol{Z}_{(j)}^\top(\boldsymbol{Y}_{(j)} - \boldsymbol{X}_{(j)}\boldsymbol{\beta})/\sigma^2$. To evaluate this integral, we used the pseudo-adaptive quadrature rule proposed by Rizopoulos (2012b), integrating over $\boldsymbol{\alpha} = 2^{-1/2}\boldsymbol{B}_{b,(j)}(\boldsymbol{b}-\boldsymbol{\mu}_{b,(j)})$ instead of $\boldsymbol{b}$, where $\boldsymbol{B}_{b,(j)}^\top\boldsymbol{B}_{b,(j)} = \boldsymbol{\Sigma}_{b,(j)}^{-1}$. By noting that $\boldsymbol{b}$ is equal to $\boldsymbol{\mu}_{b,(j)}+\sqrt{2}\boldsymbol{B}_{b,(j)}^{-1}\boldsymbol{\alpha}$, the Jacobian of the transformation is $|\frac{\partial\boldsymbol{b}}{\partial\boldsymbol{\alpha}^\top}| = 2^{q/2}|\boldsymbol{B}_{b,(j)}|^{-1}$. Note also that, as $\boldsymbol{\Sigma}_{b,(j)}$ is positive definite, the entries in the main diagonal of $\boldsymbol{B}_{b,(j)}$ are positive, and thus the Jacobian is also positive. Since $(\boldsymbol{b}-\boldsymbol{\mu}_{b,(j)})^\top\boldsymbol{\Sigma}_{b,(j)}^{-1}(\boldsymbol{b}-\boldsymbol{\mu}_{b,(j)})/2 = \boldsymbol{\alpha}^\top\boldsymbol{\alpha}$, it easily follows that the dropout contribution can be written as

$$f(T,\delta|\boldsymbol{Y}_{(j)};\boldsymbol{\theta}) = \pi^{-q/2}\int f(T,\delta|\boldsymbol{b} = \boldsymbol{\mu}_{b,(j)} + \sqrt{2}\boldsymbol{B}_{b,(j)}^{-1}\boldsymbol{\alpha};\boldsymbol{\theta})e^{-\boldsymbol{\alpha}^\top\boldsymbol{\alpha}}d\boldsymbol{\alpha},$$

which can be approximated by Gauss-Hermite quadrature using the formula

$$f(T,\delta|\boldsymbol{Y}_{(j)};\boldsymbol{\theta}) \simeq \pi^{-q/2}\sum_{i_1=1}^{nQH}\ldots\sum_{i_q=1}^{nQH}\omega_{i_1}\ldots\omega_{i_q}f\left\{T,\delta|\boldsymbol{b} = \boldsymbol{\mu}_{b,(j)} + 2^{1/2}\boldsymbol{B}_{b,(j)}^{-1}\begin{pmatrix}x_{i_1}\\ \vdots \\ x_{i_q}\end{pmatrix};\boldsymbol{\theta}\right\},$$

where $x_1,\ldots,\omega_{nQH}$ and $\omega_1,\ldots\omega_{nQH}$ denote the nQH-point Gauss-Hermite product rule abscissas and weights, respectively. To calculate the asymptotic bias in the joint model, though, we need to calculate the expectations, $E_{\boldsymbol{Y}_{(j)}|M=j;\boldsymbol{\theta}_0}\left\{\log f(T,\delta|\boldsymbol{Y}_{(j)};\boldsymbol{\theta})|M = j;\boldsymbol{\theta}_0\right\}$, with respect to the true mechanism generating the data $(\boldsymbol{Y}_{(M)}, M)$. These are equal

to

$$-\frac{q}{2}\log(\pi)+\int \log\left[\int \left\{f(T,\delta|\boldsymbol{b}=\boldsymbol{\mu}_{b,(j)}+\sqrt{2}\boldsymbol{B}_{b,(j)}^{-1}\boldsymbol{\alpha};\boldsymbol{\theta})e^{-\boldsymbol{\alpha}^\top\boldsymbol{\alpha}}\right\}d\boldsymbol{\alpha}\right] f_0(\boldsymbol{Y}_{(j)}|M=j;\boldsymbol{\theta}_0)d\boldsymbol{Y}_{(j)}.$$

The inner integral was evaluated using Gauss-Hermite quadrature with 15 points as shown above. For the deterministic MAR dropout scenario, the outer integral was approximated through quasi-Monte Carlo integration as the one used in Subsection 2.1, whereas for the stochastic MAR dropout scenario, through quasi-Monte Carlo coupled with importance sampling as the one described in Subsection 3.2.3. We used 10^4 quasi-Monte Carlo draws. To solve the equations required, we used the BFGS algorithm (Thisted, 1988), which does not require the explicit evaluation of the Hessian matrix. The first-order partial derivatives were approximated using the forward difference approximation.

We used the same data generating mechanisms as those described in Section 3.2. The results are presented graphically on the right hand side of Figure 3.1. The implications remain the same; the bias in the slope estimate increases with higher MAR dropout probability, though the bias in the PH-SREM(CV) model is lower than the bias in the LN-SREM(RE) model.

Both the PH-SREM(CV) and LN-SREM(RE) models assume continuous dropout time, as it is the case in most real data examples. However, mainly for computational reasons, i.e. to avoid integration over the dropout times, the "true" dropout time was assumed to be discrete. To investigate whether part of the calculated bias in the slope estimate was due to this misspecification, we also evaluated the bias in a SREM(CV) model assuming discrete dropout time. The SREM(CV) model was of the form, $P(M=j|M\geq j,\boldsymbol{b})=\frac{\exp\{l_1+l_2m(t_j)\}}{1+\exp\{l_1+l_2m(t_j)\}}$ for $j<Q$, i.e. the hazard of $M=j$ is equal to the "true" marker value at t_j (recall that $m(t_j)=(\beta_0+b_0)+(\beta_1+b_1)t_j$ is the "true" marker value at t_j), and $P(M=j|M\geq j,\boldsymbol{b})=1$, for $j=Q$. The bias calculation was performed in the same way as with the continuous-time PH-SREM(CV) model, after replacing $f(T,\delta|\boldsymbol{Y}_{(j)};\boldsymbol{\theta})$ with

$$\prod_{k=1}^{j}\left[1+\exp\left\{l_1+l_2m(t_k)\right\}\right]^{-1}\exp\left\{l_1+l_2m(t_j)\right\},\ \text{for } j<Q,$$

and $\prod_{k=1}^{j-1}\left[1+\exp\left\{l_1+l_2m(t_k)\right\}\right]^{-1}$ for $j=Q$. We compared the bias in the estimated slope using the continuous-time PH-SREM(CV) model with natural splines for the baseline hazard function with the corresponding bias in the discrete PH-SREM(CV) using the same parameters regarding the true dropout mechanisms. The results for the discrete SREM on the current value are presented in Figure 3.2. The maximum difference between the slope estimates was quite minor (2.10%) for the stochastic true dropout model, and small (-10.28%) for the deterministic true dropout model. It should be also emphasized that, under the deterministic true dropout model, the bias in the estimated slope was higher in the discrete SREM(CV) compared to the continuous PH-SREM(CV) model. Thus, these results suggest that evaluating a continuous-time SREM under a discrete dropout mechanism does not seem to have strong influence on the estimate of the population slope. Also, our findings on the increasing bias in the estimated slope, as the MAR dropout probability becomes higher, still hold.

3.4 Proposed Model

3.4.1 Structure of the proposed model

A major assumption made by most SREMs is that the marker measurements (for example modeled by an LMM) and the time to dropout are independent given the random effects. This further implies that, in these models, the dropout mechanism is allowed to be either MNAR or MCAR; i.e. MAR cannot hold without reducing to MCAR (Njagi et al. 2014). Under a MAR true dropout mechanism, the dropout probabilities by definition depend on the observed data. In this sense, the estimated association parameters of an SREM applied to MAR data are likely to differ from zero, which in turn may affect the estimates of the marker population parameters (i.e. the slope estimate).

Motivated by the definition of MNAR, which implies dependence on the missing observations after conditioning on the observed ones in at least one missingness pattern and one realisation, we propose an SREM model in which the hazard of dropout

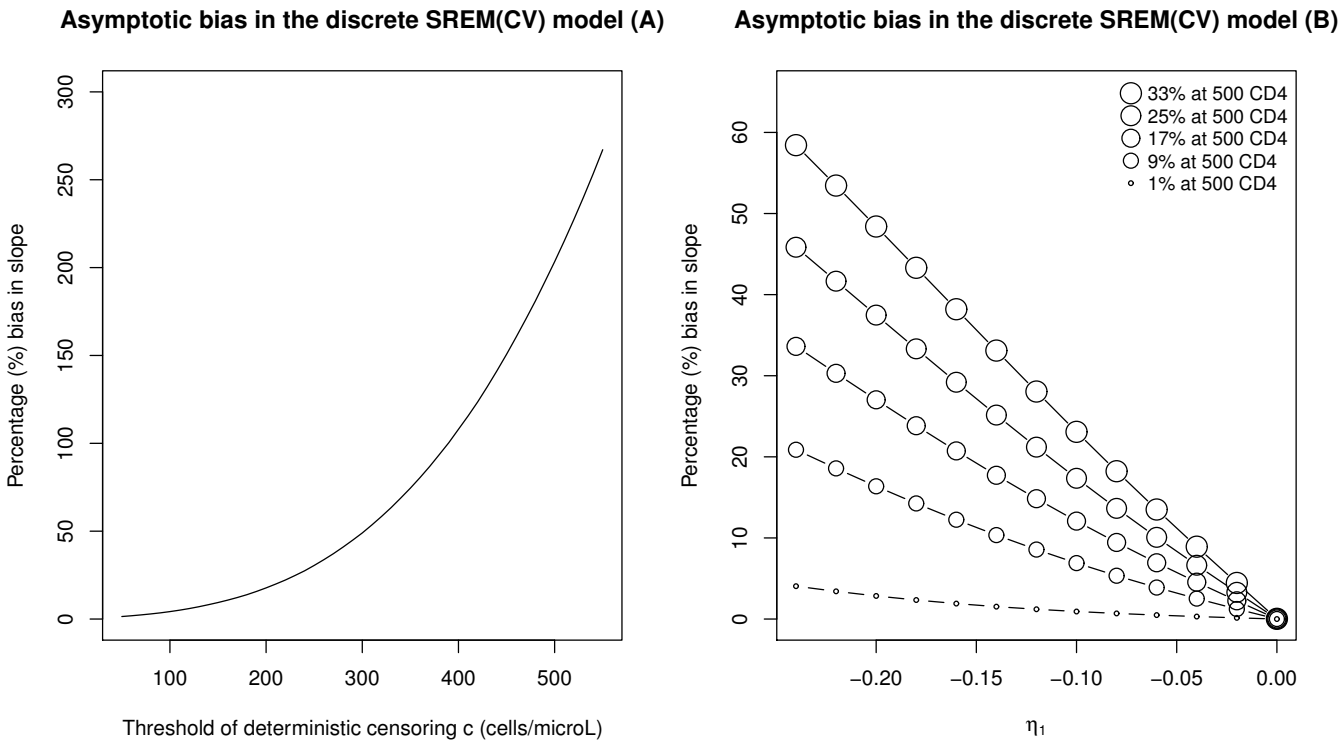


Figure 3.2: Asymptotic bias in a disease's marker rate of change estimated by a discrete SREM model using the current-value parameterization assuming MAR dropout mechanisms: subjects drop out when the marker reaches a certain threshold c (A; deterministic dropout) or when the probability of dropout is a function of marker values (B; stochastic dropout). In the latter case, η_1 measures the change in the log-odds of dropout associated with one unit decrease in the current marker value. The size of the plotting symbol is proportional to the dropout probability at 500 CD4 cells/μL, i.e. approximately the CD4 counts at seroconversion (baseline).

depends on the last observed marker value, as well as on the random effects, referred to as PH-SREM(LV,RE) from now on. That is, the marker enters as an explanatory covariate in the dropout process, which, in contrast to most SREMs, implies that the marker and dropout processes are not independent given the random effects. To distinguish subjects, we reintroduce subscript i assuming also that the maximum number of observations and the visit times may differ between subjects. Given $M_i = m_i$, the observed marker measurements on subject i are $\boldsymbol{Y}_{i,(m_i)}^{\top} = (Y_{i1}, \ldots, Y_{im_i})$ collected at times $t_{i1}, \ldots, t_{im_i}$. We also assume that the observed dropout time T_i and its associated dropout indicator δ_i are exactly known for each subject, as is the case in most applications. If $t_{ij} \leq t < t_{ij+1}$, the assumed hazard function of the model is

$$h\{t|\boldsymbol{b}_i, \mathcal{Y}_i(t); \boldsymbol{\theta}\} = h_0(t; \boldsymbol{\psi}) \exp\left\{\boldsymbol{\gamma}^{\top}\boldsymbol{w}_i + \boldsymbol{\alpha}^{\top}\boldsymbol{b}_i + \boldsymbol{\phi}^{\top}\boldsymbol{g}(y_{ij})\right\}, \tag{3.33}$$

where $\boldsymbol{g}$ is a potentially vector-valued function of the most recent observed marker values, associated with a parameter $\boldsymbol{\phi}$, and the parameter $\boldsymbol{\alpha}$ measures the association of the dropout model with the random effects, after taking into account the last value observed. Also, $h_0(t)$ is the baseline hazard function, assumed to be constant within each of J pre-specified intervals, $\boldsymbol{w}_i$ is a vector of baseline covariates with a corresponding vector of regression parameters $\boldsymbol{\gamma}$ and $\mathcal{Y}_i(t)$ denotes the history of observed marker values up to time point t. When the association parameter $\boldsymbol{\alpha}$ is different from zero, a MNAR missingness mechanism is implied, whereas if $\boldsymbol{\alpha} = \boldsymbol{0}$, the model suggests that missingness is MAR or MCAR, depending on the value of $\boldsymbol{\phi}$ (i.e. $\boldsymbol{\alpha} = \boldsymbol{0}$ with $\boldsymbol{\phi} = \boldsymbol{0}$ implies MCAR, while $\boldsymbol{\alpha} = \boldsymbol{0}$ with $\boldsymbol{\phi} \neq \boldsymbol{0}$ implies MAR dropout). The implied likelihood contribution for the ith subject is $f(\boldsymbol{Y}_{i,(m_i)}, T_i, \delta_i; \boldsymbol{\theta}) = \int f(\boldsymbol{Y}_{i,(m_i)}|\boldsymbol{b}_i; \boldsymbol{\theta}) f\{T_i, \delta_i|\mathcal{Y}_i(T_i), \boldsymbol{b}_i; \boldsymbol{\theta}\} f(\boldsymbol{b}_i; \boldsymbol{\theta}) d\boldsymbol{b}_i$, where the time-to-event contribution is equal to

$$\begin{aligned} f\{T_i, \delta_i|\mathcal{Y}_i(T_i), \boldsymbol{b}_i; \boldsymbol{\theta}\} &= \left\{h_0(T_i; \boldsymbol{\psi}) e^{\boldsymbol{\gamma}^{\top}\boldsymbol{\omega}_i + \boldsymbol{\alpha}^{\top}\boldsymbol{b}_i + \boldsymbol{\phi}^{\top}\boldsymbol{g}(Y_{im_i})}\right\}^{\delta_i} \\ &\times \exp\left\{ - \sum_{k=1}^{m_i - 1} \int_{t_{ik}}^{t_{ik+1}} h_0(u; \boldsymbol{\psi}) e^{\boldsymbol{\gamma}^{\top}\boldsymbol{\omega}_i + \boldsymbol{\alpha}^{\top}\boldsymbol{b}_i + \boldsymbol{\phi}^{\top}\boldsymbol{g}(Y_{ik})} du \right. \\ &\left. - \int_{t_{im_i}}^{T_i} h_0(u; \boldsymbol{\psi}) e^{\boldsymbol{\gamma}^{\top}\boldsymbol{\omega}_i + \boldsymbol{\alpha}^{\top}\boldsymbol{b}_i + \boldsymbol{\phi}^{\top}\boldsymbol{g}(Y_{im_i})} du \right\}. \end{aligned} \tag{3.34}$$

For the marker model we assume a standard LMM of the form $\boldsymbol{Y}_{i,(m_i)} = \boldsymbol{X}_{i,(m_i)}\boldsymbol{\beta} + \boldsymbol{Z}_{i,(m_i)}\boldsymbol{b}_i + \boldsymbol{\epsilon}_{i,(m_i)}$, where $\boldsymbol{b}_i \sim N(\boldsymbol{0}, \boldsymbol{D})$ and $\boldsymbol{\epsilon}_{i,(m_i)} \sim N(\boldsymbol{0}, \omega^{-1}\boldsymbol{I}_{(m_i)})$.

3.4.2 Outline of the MCMC algorithm

To estimate the model parameters, we developed a Bayesian inferential procedure based on a Markov chain Monte Carlo (MCMC) method. A normal prior distribution, $N(\boldsymbol{\mu}_0, \boldsymbol{C}_0)$, was used for $\boldsymbol{\beta}$, a Gamma(τ_1, τ_2) distribution with shape τ_1 and rate τ_2, $f(\omega) \propto \omega^{\tau_1 - 1}e^{-\tau_2\omega}$, for ω and a normal, $N(\boldsymbol{\mu}_0^s, \boldsymbol{C}_0{}^s)$, distribution for the parameter vector of the dropout model, $\boldsymbol{\beta}^{s\top} = (\boldsymbol{\gamma}^\top, \boldsymbol{\phi}^\top, \boldsymbol{\alpha}^\top)$. For the covariance matrix of the random effects, $\boldsymbol{D}$, we assumed the Inverse-Wishart distribution, i.e. $\boldsymbol{D} \sim \text{InvWish}(\boldsymbol{A}, df)$. Independent Gamma$(a_{0k}, \lambda_{0k})$ prior distributions, $k = 1, \ldots, J$, were assumed for the baseline hazard parameters. Letting $\mathcal{D} = \{(\boldsymbol{Y}_{i,(m_i)}, \boldsymbol{X}_{i,(m_i)}, \boldsymbol{Z}_{i,(m_i)}, \boldsymbol{\omega}_i, T_i, \delta_i); i = 1, 2, \ldots, N\}$ be the observed data, the posterior distribution of the parameters is proportional to

$$
\begin{aligned}
f(\boldsymbol{\theta}, \{\boldsymbol{b}_i\}_{i=1}^N | \mathcal{D}) \quad &\propto \quad f(\boldsymbol{\theta}) \prod_{i=1}^{N} \Bigg[\omega^{m_i/2} \exp\left\{ -\frac{\omega}{2} \| \boldsymbol{Y}_{i,(m_i)} - \boldsymbol{X}_{i,(m_i)}\boldsymbol{\beta} - \boldsymbol{Z}_{i,(m_i)}\boldsymbol{b}_i \|^2 \right\} \\
&\times \quad |\boldsymbol{D}|^{-1/2} \exp\left\{ -\frac{1}{2} \boldsymbol{b}_i^\top \boldsymbol{D}^{-1} \boldsymbol{b}_i \right\} f\{T_i, \delta_i | \mathcal{Y}_i(T_i), \boldsymbol{b}_i; \boldsymbol{\theta}\} \Bigg].
\end{aligned}
$$

Given the likelihood and prior specifications, it immediately follows that the conditional posterior distribution of $\boldsymbol{\beta}$ given the remaining parameters is multivariate normal with covariance matrix $\boldsymbol{C}_1 = (\boldsymbol{C}_0^{-1} + \omega \sum_{i=1}^N \boldsymbol{X}_{i,(m_i)}^\top \boldsymbol{X}_{i,(m_i)})^{-1}$ and mean $\boldsymbol{\mu}_1 = \boldsymbol{C}_1\{\boldsymbol{C}_0^{-1}\boldsymbol{\mu}_0 + \omega \sum_{i=1}^N \boldsymbol{X}_{i,(m_i)}^\top (\boldsymbol{Y}_{i,(m_i)} - \boldsymbol{Z}_{i,(m_i)}\boldsymbol{b}_i)\}$, the conditional posterior distribution of ω is the Gamma$(n/2 + \tau_1, \tau_2 + \sum_{i=1}^N \| \boldsymbol{Y}_{i,(m_i)} - \boldsymbol{X}_{i,(m_i)}\boldsymbol{\beta} - \boldsymbol{Z}_{i,(m_i)}\boldsymbol{b}_i \|^2/2)$ distribution, where $n = \sum_{i=1}^N m_i$, and the conditional distribution of $\boldsymbol{D}$ is the InvWish$(\boldsymbol{A} + \sum_{i=1}^N \boldsymbol{b}_i\boldsymbol{b}_i^\top, df + N)$ distribution.

3.4.2.1 Conditional posterior distribution of the random effects, $\boldsymbol{b}_i$

To efficiently simulate random draws from the conditional distribution of $\boldsymbol{b}_i$, we used rejection sampling. The idea behind rejection sampling is to first simulate from a dif-

ferent distribution (the proposal distribution) and then accept the proposed value with a specified probability. If the probability of acceptance fulfils a specific condition, the accepted values will constitute a sample from the target distribution. After dropping the dependence on the observed data and the remaining parameters for notational convenience, the conditional posterior distribution of $\boldsymbol{b}_i$, which is our target distribution, is proportional to

$$\begin{aligned} f(\boldsymbol{b}_i) &\propto \exp\Big\{ -\frac{1}{2}\boldsymbol{b}_i^\top(\boldsymbol{D}^{-1}+\omega\boldsymbol{Z}_{i,(m_i)}^\top\boldsymbol{Z}_{i,(m_i)})\boldsymbol{b}_i \\ &+ \omega\boldsymbol{b}_i^\top\boldsymbol{Z}_{i,(m_i)}^\top(\boldsymbol{Y}_{i,(m_i)}-\boldsymbol{X}_{i,(m_i)}\boldsymbol{\beta}) + \delta_i\boldsymbol{b}_i^\top\boldsymbol{\alpha} - e^{\boldsymbol{b}_i^\top\boldsymbol{\alpha}}q_i\Big\}, \end{aligned}$$

where q_i is a positive constant with respect to $\boldsymbol{b}_i$. Note that we have suppressed the dependence on the data and on the other parameters to simplify the notation. Our proposal distribution requires the posterior mode. This can be calculated by using the Newton-Raphson (NR) algorithm working on the log scale. The vector of the first-order partial derivatives is

$$\mathcal{U}(\boldsymbol{b}_i) = -(\boldsymbol{D}^{-1}+\omega\boldsymbol{Z}_{i,(m_i)}^\top\boldsymbol{Z}_{i,(m_i)})\boldsymbol{b}_i + \omega\boldsymbol{Z}_{i,(m_i)}^\top(\boldsymbol{Y}_{i,(m_i)}-\boldsymbol{X}_{i,(m_i)}\boldsymbol{\beta}) + \delta_i\boldsymbol{\alpha} - \boldsymbol{\alpha}e^{\boldsymbol{b}_i^\top\boldsymbol{\alpha}}q_i,$$

and the information matrix is

$$\mathcal{I}(\boldsymbol{b}_i) = (\boldsymbol{D}^{-1}+\omega\boldsymbol{Z}_{i,(m_i)}^\top\boldsymbol{Z}_{i,(m_i)}) + \boldsymbol{\alpha}\boldsymbol{\alpha}^\top e^{\boldsymbol{b}_i^\top\boldsymbol{\alpha}}q_i.$$

Letting $\widehat{\boldsymbol{b}}_i^{(t)}$ be the current estimate of the posterior mode, the next estimate is obtained by the formula $\widehat{\boldsymbol{b}}_i^{(t+1)} = \widehat{\boldsymbol{b}}_i^{(t)} + \mathcal{I}(\widehat{\boldsymbol{b}}_i^{(t)})^{-1}\mathcal{U}(\widehat{\boldsymbol{b}}_i^{(t)})$. After very few NR iterations (about 2 or 3 in our examples) we get the posterior mode $\widehat{\boldsymbol{b}}_i$. We take the proposal distribution to be a multivariate normal density with mean $\widehat{\boldsymbol{b}}_i$ and covariance matrix $(\boldsymbol{D}^{-1}+\omega\boldsymbol{Z}_{i,(m_i)}^\top\boldsymbol{Z}_{i,(m_i)})^{-1}$. Thus,

$$g(\boldsymbol{b}_i) \propto \exp\left\{-\frac{1}{2}\boldsymbol{b}_i^\top(\boldsymbol{D}^{-1}+\omega\boldsymbol{Z}_{i,(m_i)}^\top\boldsymbol{Z}_{i,(m_i)})\boldsymbol{b}_i + \boldsymbol{b}_i^\top(\boldsymbol{D}^{-1}+\omega\boldsymbol{Z}_{i,(m_i)}^\top\boldsymbol{Z}_{i,(m_i)})\widehat{\boldsymbol{b}}_i\right\}.$$

To apply the rejection algorithm, we need to ensure that there exists a $K > 0$ such that $f(\boldsymbol{b}_i) \leq Kg(\boldsymbol{b}_i)$, for all $\boldsymbol{b}_i$. Thus, $K \geq f(\boldsymbol{b}_i)/g(\boldsymbol{b}_i)$, which means that $K = \max\{f(\boldsymbol{b}_i)/g(\boldsymbol{b}_i)\}$. The ratio of $f(\boldsymbol{b}_i)$ to $g(\boldsymbol{b}_i)$ on the log scale is equal to

$$\ell(\boldsymbol{b}_i) = \omega\boldsymbol{b}_i^\top\boldsymbol{Z}_{i,(m_i)}^\top(\boldsymbol{Y}_{i,(m_i)}-\boldsymbol{X}_{i,(m_i)}\boldsymbol{\beta}) + \delta_i\boldsymbol{b}_i^\top\boldsymbol{\alpha} - e^{\boldsymbol{b}_i^\top\boldsymbol{\alpha}}q_i - \boldsymbol{b}_i^\top(\boldsymbol{D}^{-1}+\omega\boldsymbol{Z}_{i,(m_i)}^\top\boldsymbol{Z}_{i,(m_i)})\widehat{\boldsymbol{b}}_i,$$

after dropping the additive constants. Since $\widehat{\boldsymbol{b}}_i$ maximizes $f(\boldsymbol{b}_i)$, it must be the case that $\mathcal{U}(\widehat{\boldsymbol{b}}_i) = \mathbf{0}$, which implies that $-(\boldsymbol{D}^{-1} + \omega \boldsymbol{Z}_{i,(m_i)}^\top \boldsymbol{Z}_{i,(m_i)})\widehat{\boldsymbol{b}}_i + \omega \boldsymbol{Z}_{i,(m_i)}^\top (\boldsymbol{Y}_{i,(m_i)} - \boldsymbol{X}_{i,(m_i)}\boldsymbol{\beta}) + \delta_i \boldsymbol{\alpha} = \boldsymbol{\alpha} e^{\widehat{\boldsymbol{b}}_i^\top \boldsymbol{\alpha}} q_i$. Thus,

$$\ell(\boldsymbol{b}_i) = \boldsymbol{b}_i^\top \boldsymbol{\alpha} e^{\widehat{\boldsymbol{b}}_i^\top \boldsymbol{\alpha}} q_i - e^{\boldsymbol{b}_i^\top \boldsymbol{\alpha}} q_i.$$

Note that $\ell(\boldsymbol{b}_i)$ depends on $\boldsymbol{b}_i$ only through the term $\boldsymbol{b}_i^\top \boldsymbol{\alpha}$ and it does not depend on the individual elements of $\boldsymbol{b}_i$. Therefore, we can make the change of variables $x_i = \boldsymbol{b}_i^\top \boldsymbol{\alpha}$, leading to $\ell(x_i) = x_i e^{\widehat{x}_i} q_i - e^{x_i} q_i$, where $\widehat{x}_i = \widehat{\boldsymbol{b}}_i^\top \boldsymbol{\alpha}$. This means that $L'(x_i) = e^{\widehat{x}_i} q_i - e^{x_i} q_i$, and thus $L'(x_i) = 0 \Rightarrow x_i = \widehat{x}_i$. To show that $\widehat{x}_i$ is indeed a maximum, we need to ensure that the second derivative is negative, which immediately follows as $L''(x_i) = -e^{x_i} q_i < 0$. Thus, $K = \exp\left(\widehat{\boldsymbol{b}}_i^\top \boldsymbol{\alpha} e^{\widehat{\boldsymbol{b}}_i^\top \boldsymbol{\alpha}} q_i - e^{\widehat{\boldsymbol{b}}_i^\top \boldsymbol{\alpha}} q_i\right)$. To carry out the rejection algorithm, we first simulate a value $\boldsymbol{b}_i^{can}$ from $g(\boldsymbol{b}_i)$ and then accept the proposed value if $u \leq \frac{f(\boldsymbol{b}_i^{can})}{K g(\boldsymbol{b}_i^{can})}$, that is, if

$$u \leq \exp\left\{ (\boldsymbol{b}_i^{can\top} \boldsymbol{\alpha} - \widehat{\boldsymbol{b}}_i^\top \boldsymbol{\alpha}) e^{\widehat{\boldsymbol{b}}_i^\top \boldsymbol{\alpha}} q_i - (e^{\boldsymbol{b}_i^{can\top} \boldsymbol{\alpha}} - e^{\widehat{\boldsymbol{b}}_i^\top \boldsymbol{\alpha}}) q_i \right\},$$

with u being a random value from the uniform distribution over the interval $(0, 1)$.

3.4.2.2 Conditional posterior distribution of the parameters of the dropout mechanism, $(\boldsymbol{\beta}^s, \boldsymbol{\psi})$

Finally, the parameters of the dropout mechanism are updated via a Metropolis-Hastings step described below. Let $\boldsymbol{X}_{i,(m_i)}^s$ be a matrix with the j-th row being $\boldsymbol{x}_{ij}^{s\top} = (\boldsymbol{\omega}_i^\top, \boldsymbol{g}(y_{ij})^\top, \boldsymbol{b}_i^\top)$, δ_{ij} be the failure indicator on the jth occasion, $j = 1, \ldots, m_i$, $\boldsymbol{t}_i^0 = (t_{i1}, \ldots, t_{im_i})^\top$ and $\boldsymbol{t}_i^1 = (t_{i2}, \ldots, t_{im_i - 1}, T_i)^\top$. Note that $\delta_{ij} = 0$, for $j < m_i$, and $\delta_{ij} = \delta_i$, for $j = m_i$. It follows that

$$\begin{aligned} f(T_i, \delta_i | \mathcal{Y}_i(T_i), \boldsymbol{b}_i; \boldsymbol{\beta}^s, \boldsymbol{\psi}) &= \prod_{j=1}^{m_i} \left\{ h_0(t_{ij}^1; \boldsymbol{\psi}) \exp\left(\boldsymbol{x}_{ij}^{s\top} \boldsymbol{\beta}^s\right) \right\}^{\delta_{ij}} \\ \times \quad \exp\left\{ -H_0(t_{ij}^1; \boldsymbol{\psi}) e^{\boldsymbol{x}_{ij}^{s\top} \boldsymbol{\beta}^s} + H_0(t_{ij}^0; \boldsymbol{\psi}) e^{\boldsymbol{x}_{ij}^{s\top} \boldsymbol{\beta}^s} \right\}, \end{aligned} \tag{3.35}$$

since $\int_{t_{ij}^0}^{t_{ij}^1} h_0(u; \boldsymbol{\psi}) du = H_0(t_{ij}^1; \boldsymbol{\psi}) - H_0(t_{ij}^0; \boldsymbol{\psi})$. Since we have assumed that the baseline hazard function remains constant within the pre-specified intervals $[s_{g-1}, s_g)$, it

follows that $H_0(t;\boldsymbol{\psi}) = \sum_{k=1}^{J} I_k(t)\left\{\psi_k(t - s_{k-1}) + \sum_{g=1}^{k-1}\psi_g(s_g - s_{g-1})\right\}$, with $I_k(t) = I(s_{k-1} \leq t < s_k)$. Thus, from Equation (3.35), the conditional likelihood of the dropout model is (Ibrahim et al., 2001, Sahu et al., 1997)

$$\prod_{i=1}^{N} f(T_i, \delta_i|\mathcal{Y}_i(T_i), \boldsymbol{b}_i; \boldsymbol{\beta}^s, \boldsymbol{\psi}) = \prod_{i=1}^{N}\prod_{j=1}^{m_i}\prod_{k=1}^{J}\left\{\psi_k e^{\boldsymbol{x}_{ij}^{s\top}\boldsymbol{\beta}^s}\right\}^{I_k(t_{ij}^1)\delta_i} \tag{3.36}$$

$$\exp\left[- I_k(t_{ij}^1)\left\{\psi_k(t_{ij}^1 - s_{k-1}) + \sum_{g=1}^{k-1}\psi_g(s_g - s_{g-1})\right\} e^{\boldsymbol{x}_{ij}^{s\top}\boldsymbol{\beta}^s}\right.$$

$$\left. + I_k(t_{ij}^0)\left\{\psi_k(t_{ij}^0 - s_{k-1}) + \sum_{g=1}^{k-1}\psi_g(s_g - s_{g-1})\right\} e^{\boldsymbol{x}_{ij}^{s\top}\boldsymbol{\beta}^s}\right].$$

However, it is more convenient to write down the likelihood as a product of independent contributions over the J intervals. Note that the first term in Equation (3.36) involves only the set of patients failing in the k-th interval, which is denoted by $\mathcal{D}_k$. Thus, the likelihood contribution implied is $\psi_k^{D_k}\exp\left\{\sum_{(i,j)\in\mathcal{D}_k}\boldsymbol{x}_{ij}^{s\top}\boldsymbol{\beta}^s\right\}$, where D_k is the number of patients in $\mathcal{D}_k$. By an analogous argument, the second row of Equation (3.36) can be written as $\psi_k\sum_{(i,j)\in\mathcal{R}_k^1}\Delta_{ijk}^1 e^{\boldsymbol{x}_{ij}^{s\top}\boldsymbol{\beta}^s}$, where $\Delta_{ijk}^1 = \min\left\{t_{ij}^1, s_k\right\} - s_{k-1}$ (with $\sum_{(i,j)\in\mathcal{R}_k^1}\Delta_{ijk}^1$ denoting the total person years lived in the k-th interval by all subjects), and $\mathcal{R}_k^1 = \left\{(i,j) : t_{ij}^1 \geq s_{k-1}\right\}$, whereas the third row can be written as $\psi_k\sum_{(i,j)\in\mathcal{R}_k^0}\Delta_{ijk}^0 e^{\boldsymbol{x}_{ij}^{s\top}\boldsymbol{\beta}^s}$, where $\Delta_{ijk}^0 = \min\left\{t_{ij}^0, s_k\right\} - s_{k-1}$ and $\mathcal{R}_k^0 = \left\{(i,j) : t_{ij}^0 \geq s_{k-1}\right\}$. With this notation, it follows that

$$\prod_{i=1}^{N} f(T_i, \delta_i|\mathcal{Y}_i(T_i), \boldsymbol{b}_i; \boldsymbol{\beta}^s, \boldsymbol{\psi}) = \prod_{k=1}^{J}\psi_k^{D_k}\exp\left\{\sum_{(i,j)\in\mathcal{D}_k}\boldsymbol{x}_{ij}^{s\top}\boldsymbol{\beta}^s\right.$$

$$\left. - \psi_k\left(\sum_{(i,j)\in\mathcal{R}_k^1}\Delta_{ijk}^1 e^{\boldsymbol{x}_{ij}^{s\top}\boldsymbol{\beta}^s} - \sum_{(i,j)\in\mathcal{R}_k^0}\Delta_{ijk}^0 e^{\boldsymbol{x}_{ij}^{s\top}\boldsymbol{\beta}^s}\right)\right\}.$$

Assuming independent Gamma$(\alpha_{0k}, \lambda_{0h})$ prior distributions for ψ_k, $k = 1, 2, \ldots, J$, the conditional posterior distribution of ψ_k is a Gamma distribution with shape $D_k + \alpha_{0k}$ and rate $\lambda_{0k} + S_k$, with $S_k = \sum_{(i,j)\in\mathcal{R}_k^1}\Delta_{ijk}^1 e^{\boldsymbol{x}_{ij}^{s\top}\boldsymbol{\beta}^s} - \sum_{(i,j)\in\mathcal{R}_k^0}\Delta_{ijk}^0 e^{\boldsymbol{x}_{ij}^{s\top}\boldsymbol{\beta}^s}$. By completing the density of the Gamma$(D_k + \alpha_{0k}, \lambda_{0k} + S_k)$ distribution, we can easily integrate out the baseline hazard parameters, λ_k, resulting in the marginal posterior

distribution of $\boldsymbol{\beta}^s$

$$\prod_{k=1}^{J} \frac{f(\boldsymbol{\beta}^s) \exp\left(\sum_{(i,j)\in\mathcal{D}_k} \boldsymbol{x}_{ij}^{s\top} \boldsymbol{\beta}^s\right)}{\left(\sum_{(i,j)\in\mathcal{R}_k^1} \Delta_{ijk}^1 e^{\boldsymbol{x}_{ij}^{s\top}\boldsymbol{\beta}^s} - \sum_{(i,j)\in\mathcal{R}_k^0} \Delta_{ijk}^0 e^{\boldsymbol{x}_{ij}^{s\top}\boldsymbol{\beta}^s}\right)^{D_k+\alpha_{0k}}}.$$

Using straightforward matrix differentiation, the vector of first-order partial derivatives is $\mathcal{U}(\boldsymbol{\beta}^s) = \sum_{k=1}^{J} \left\{\sum_{(i,j)\in\mathcal{D}_k} \boldsymbol{x}_{ij}^s - (D_k + \alpha_{0k})\overline{\boldsymbol{x}}_k\right\} - \boldsymbol{C}_0^{s-1}(\boldsymbol{\beta}^s - \boldsymbol{\mu}_0^s)$, where $\overline{\boldsymbol{x}}_k$ is a weighted mean of the covariates over all subjects in the k-th risk set, i.e. $\overline{\boldsymbol{x}}_k = (\sum_{(i,j)\in\mathcal{R}_k^1} \boldsymbol{x}_{ij}^s \Delta_{ijk}^1 e^{\boldsymbol{x}_{ij}^{s\top}\boldsymbol{\beta}^s} - \sum_{(i,j)\in\mathcal{R}_k^0} \boldsymbol{x}_{ij}^s \Delta_{ijk}^0 e^{\boldsymbol{x}_{ij}^{s\top}\boldsymbol{\beta}^s})/S_k$ and the information matrix

$$\mathcal{I}(\boldsymbol{\beta}^s) = \sum_{k=1}^{J} (D_k + \alpha_{0k}) \Bigg\{ \Bigg(\sum_{(i,j)\in\mathcal{R}_k^1} \boldsymbol{x}_{ij}^s \boldsymbol{x}_{ij}^{s\top} \Delta_{ijk}^1 e^{\boldsymbol{x}_{ij}^{s\top}\boldsymbol{\beta}^s} - \sum_{(i,j)\in\mathcal{R}_k^0} \boldsymbol{x}_{ij}^s \boldsymbol{x}_{ij}^{s\top} \Delta_{ijk}^0 e^{\boldsymbol{x}_{ij}^{s\top}\boldsymbol{\beta}^s} \Bigg) / S_k - \overline{\boldsymbol{x}}_k \overline{\boldsymbol{x}}_k^\top \Bigg\} + \boldsymbol{C}_0^{s-1}.$$

Let $(\boldsymbol{\beta}^s, \boldsymbol{\psi})$ be the current value of the chain. We propose a value $\boldsymbol{\beta}_{can}^s$ from a multivariate normal distribution with mean obtained from one NR step starting from the current value and covariance matrix $\mathcal{I}(\boldsymbol{\beta}^s)^{-1}$. Given the proposed value $\boldsymbol{\beta}_{can}^s$, we propose a value $\boldsymbol{\psi}_{can}$ for $\boldsymbol{\psi}$ through its conditional posterior distribution. Then the Metropolis–Hastings acceptance probability only depends on the marginal posterior distribution of $\boldsymbol{\beta}^s$.

3.4.3 Extending the Proposed Model to the Bivariate Case

The univariate marker model can be easily extended to a bivariate one, in which two correlated longitudinally measured markers can be simultaneously modeled. Let $\boldsymbol{Y}_{i,(m_i^c)}^c = (Y_{i1}^c, \ldots, Y_{im_i^c}^c)^\top$ and $\boldsymbol{Y}_{i,(m_i^v)}^v = (Y_{i1}^v, \ldots, Y_{im_i^v}^v)^\top$ be the sequences of measurements of the first and second marker (e.g. CD4 cell count and HIV-RNA viral load, respectively) collected at times $t_{i1}^c, \ldots, t_{im_i^c}^c$ and $t_{i1}^v, \ldots, t_{im_i^v}^v$, respectively. Let $\boldsymbol{X}_{i,(m_i^c)}^c$ and $\boldsymbol{X}_{i,(m_i^v)}^v$ be the design matrices associated with the fixed effects $\boldsymbol{\beta}^B = (\boldsymbol{\beta}^{c\top}, \boldsymbol{\beta}^{v\top})^\top$ of the two markers and $\boldsymbol{Z}_{i,(m_i^c)}^c$ and $\boldsymbol{Z}_{i,(m_i^v)}^v$ be the corresponding design matrices associated with the random effects $\boldsymbol{b}_i^B = (\boldsymbol{b}_i^{c\top}, \boldsymbol{b}_i^{v\top})^\top$. The model for the evolution of both

markers is assumed to be of the form

$$\begin{pmatrix} \boldsymbol{Y}^c_{i,(m_i^c)} \\ \boldsymbol{Y}^v_{i,(m_i^v)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}^c_{i,(m_i^c)} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{X}^v_{i,(m_i^v)} \end{pmatrix} \boldsymbol{\beta}^B + \begin{pmatrix} \boldsymbol{Z}^c_{i,(m_i^c)} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Z}^v_{i,(m_i^v)} \end{pmatrix} \boldsymbol{b}_i^B + \begin{pmatrix} \boldsymbol{\epsilon}^c_{i,(m_i^c)} \\ \boldsymbol{\epsilon}^v_{i,(m_i^v)} \end{pmatrix},$$

where $\boldsymbol{\epsilon}^c_{i,(m_i^c)} \sim N(\boldsymbol{0}, \omega_c^{-1}\boldsymbol{I}_{m_i^c})$ and $\boldsymbol{\epsilon}^v_{i,(m_i^v)} \sim N(\boldsymbol{0}, \omega_v^{-1}\boldsymbol{I}_{m_i^v})$ are the within-subject residuals for both markers, respectively, assumed independent of each other. In addition, the random effects of the two markers are assumed to jointly follow the multivariate normal distribution, i.e. $\boldsymbol{b}_i^B \sim N(\boldsymbol{0}, \boldsymbol{D}^B)$. Thus, the correlation between the two markers is induced through the correlation of the random effects. The hazard of dropout is assumed to depend on the most recent values of both markers, as well as on the random effects. The posterior distribution of all unknown quantities is thus

$$\begin{aligned} f(\boldsymbol{\theta}, \boldsymbol{b}|\mathcal{D}) \;\propto\; & f(\boldsymbol{\theta}) \prod_{i=1}^{N} \Bigg\{ \omega_c^{m_i^c/2} \omega_v^{m_i^v/2} \exp\Bigg(-\frac{\omega_c}{2} \|\boldsymbol{Y}^c_{i,(m_i^c)} - \boldsymbol{X}^c_{i,(m_i^c)}\boldsymbol{\beta}^c - \boldsymbol{Z}^c_{i,(m_i^c)}\boldsymbol{b}_i^c\|^2 \\ & - \frac{\omega_v}{2} \|\boldsymbol{Y}^v_{i,(m_i^v)} - \boldsymbol{X}^v_{i,(m_i^v)}\boldsymbol{\beta}^v - \boldsymbol{Z}^v_{i,(m_i^v)}\boldsymbol{b}_i^v\|^2 - \frac{1}{2}\boldsymbol{b}_i^\top \boldsymbol{D}^{B^{-1}} \boldsymbol{b}_i \Bigg) |\boldsymbol{D}^B|^{-1/2} \\ & \times\; f(T_i, \delta_i | \mathcal{Y}_i^c(T_i), \mathcal{Y}_i^v(T_i), \boldsymbol{b}_i; \boldsymbol{\theta}) \Bigg\}, \end{aligned}$$

where $\mathcal{Y}_i^c(t)$ and $\mathcal{Y}_i^v(t)$ denote the history of observed responses for both markers up to time t and $\boldsymbol{\theta}$ denotes the whole parameter vector. Assuming independent gamma prior distributions for the within-subject precisions, i.e. $\omega_c \sim \text{Gamma}(\tau_1^c, \tau_2^c)$ and $\omega_v \sim \text{Gamma}(\tau_1^v, \tau_2^v)$, it immediately follows that the conditional posterior distributions for ω_c and ω_v are the Gamma$(n^c/2 + \tau_1^c, \tau_2^c + \sum_{i=1}^N \|\boldsymbol{Y}^c_{i,(m_i^c)} - \boldsymbol{X}^c_{i,(m_i^c)}\boldsymbol{\beta}^c - \boldsymbol{Z}^c_{i,(m_i^c)}\boldsymbol{b}_i^c\|^2/2)$ and Gamma$(n^v/2 + \tau_1^v, \tau_2^v + \sum_{i=1}^N \|\boldsymbol{Y}^v_{i,(m_i^v)} - \boldsymbol{X}^v_{i,(m_i^v)}\boldsymbol{\beta}^v - \boldsymbol{Z}^v_{i,(m_i^v)}\boldsymbol{b}_i^v\|^2/2)$ distributions, respectively, with $n^c = \sum_{i=1}^N m_i^c$ and $n^v = \sum_{i=1}^N m_i^v$. Defining $\boldsymbol{W}_i$ to be a diagonal $(m_i^c + m_i^v) \times (m_i^c + m_i^v)$ matrix with the main diagonal taking the value ω_c for the first m_i^c entries and the value ω_v thereafter, the conditional posterior distribution of $\boldsymbol{\beta}^B$ becomes

$$\exp\left\{ -\frac{1}{2} \sum_{i=1}^{N} (\boldsymbol{Y}^B_{i,(m_i)} - \boldsymbol{X}^B_{i,(m_i)}\boldsymbol{\beta}^B - \boldsymbol{Z}^B_{i,(m_i)}\boldsymbol{b}_i^B)^\top \boldsymbol{W}_i (\boldsymbol{Y}^B_{i,(m_i)} - \boldsymbol{X}^B_{i,(m_i)}\boldsymbol{\beta}^B - \boldsymbol{Z}^B_{i,(m_i)}\boldsymbol{b}_i^B) \right\} f(\boldsymbol{\beta}^B),$$

with $\boldsymbol{X}^B_{i,(m_i)} = \begin{pmatrix} \boldsymbol{X}^c_{i,(m_i^c)} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{X}^v_{i,(m_i^v)} \end{pmatrix}$, $\boldsymbol{Z}^B_{i,(m_i)} = \begin{pmatrix} \boldsymbol{Z}^c_{i,(m_i^c)} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Z}^v_{i,(m_i^v)} \end{pmatrix}$ and $\boldsymbol{Y}^B_{i,(m_i)} = (\boldsymbol{Y}^c_{i,(m_i^c)}{}^\top, \boldsymbol{Y}^v_{i,(m_i^v)}{}^\top)^\top$. Assuming a $N(\boldsymbol{\mu}_0, \boldsymbol{C}_0)$ prior distribution for $\boldsymbol{\beta}^B$, it can be easily shown that the conditional posterior distribution of $\boldsymbol{\beta}^B$ is the $N(\boldsymbol{\mu}_1, \boldsymbol{C}_1)$ distribution, with $\boldsymbol{C}_1 = (\boldsymbol{C}_0^{-1} + \sum_{i=1}^N \boldsymbol{X}^B_{i,(m_i)}{}^\top \boldsymbol{W}_i \boldsymbol{X}^B_{i,(m_i)})^{-1}$ and $\boldsymbol{\mu}_1 = \boldsymbol{C}_1\{\boldsymbol{C}_0^{-1}\boldsymbol{\mu}_0 + \sum_{i=1}^N \boldsymbol{X}^B_{i,(m_i)}{}^\top \boldsymbol{W}_i(\boldsymbol{Y}^B_{i,(m_i)} - \boldsymbol{Z}^B_{i,(m_i)}\boldsymbol{b}_i^B)\}$. Moreover, similarly to the univariate case, assuming an Inverse-Wishart prior distribution for $\boldsymbol{D}^B$, i.e. $\boldsymbol{D}^B \sim \text{InvWish}(\boldsymbol{A}, df)$, it can be shown that the conditional posterior distribution of $\boldsymbol{D}^B$ is the InvWish$(\boldsymbol{A} + \sum_{i=1}^N \boldsymbol{b}_i^B \boldsymbol{b}_i^{B\top}, df + G)$ distribution. Assuming that the markers were measured at the same times (i.e. $t^c_{ij} = t^v_{ij} = t_{ij}$, $j = 1, \ldots, m_i$), the dropout likelihood contribution is

$$\begin{aligned} &\Big[h_0(T_i;\boldsymbol{\psi}) \exp\Big\{\boldsymbol{\gamma}^\top \boldsymbol{\omega}_i + \boldsymbol{\alpha}^{B\top} \boldsymbol{b}_i^B + \boldsymbol{\phi}^{c\top} \boldsymbol{g}^c(y^c_{im_i}) + \boldsymbol{\phi}^{v\top} \boldsymbol{g}^v(y^v_{im_i})\Big\}\Big]^{\delta_i} \\ &\times \exp\Bigg\{ - \sum_{k=1}^{m_i-1} \int_{t_{ik}}^{t_{ik+1}} h_0(u;\boldsymbol{\psi}) e^{\boldsymbol{\gamma}^\top \boldsymbol{\omega}_i + \boldsymbol{\alpha}^{B\top} \boldsymbol{b}_i^B + \boldsymbol{\phi}^{c\top} \boldsymbol{g}^c(y^c_{ik}) + \boldsymbol{\phi}^{v\top} \boldsymbol{g}^v(y^v_{ik})} du \\ &- \int_{t_{im_i}}^{T_i} h_0(u;\boldsymbol{\psi}) e^{\boldsymbol{\gamma}^\top \boldsymbol{\omega}_i + \boldsymbol{\alpha}^{B\top} \boldsymbol{b}_i^B + \boldsymbol{\phi}^{c\top} \boldsymbol{g}^c(y^c_{im_i}) + \boldsymbol{\phi}^{v\top} \boldsymbol{g}^v(y^v_{im_i})} du \Bigg\}, \end{aligned} \tag{3.37}$$

where $\boldsymbol{\phi}^c$ and $\boldsymbol{\phi}^v$ are parameters quantifying the dependence of the hazard of dropout on the most recent values of the first and second marker, respectively, and the parameter $\boldsymbol{\alpha}^B$ measures the association of the dropout time with the random effects of both markers. Assuming also that the baseline hazard function is constant within pre-specified intervals and letting $\boldsymbol{X}^s_{i,(m_i)}$ be a matrix with the jth row being $\boldsymbol{x}^s_{ij}{}^\top = (\boldsymbol{\omega}_i^\top, \boldsymbol{g}^c(y^c_{ij})^\top, \boldsymbol{g}^v(y^v_{ij})^\top, \boldsymbol{b}_i^{B\top})$,$j = 1, 2, \ldots, m_i$, we can update the parameters of the dropout model using the algorithm described in 3.4.2.2. If the two markers have not been measured at exactly the same times, let $t_{i1}, t_{i2}, \ldots, t_{iQ}$ be the the distinct ordered measurement times arising from both markers on subject i. Then we can still use the algorithm in 3.4.2.2, but we should use the last observed markers' values in the jth row of $\boldsymbol{X}^s_{i,(m_i)}$, $j = 1, 2, \ldots Q$. Thus, some data manipulation is needed in this case. In our motivating example, CD4 cell count and HIV-RNA viral load were mostly measured at exactly the same times.

Finally, the conditional posterior distribution of the random effects is proportional

to

$$f(\boldsymbol{b}_i^B) \propto \exp\Big\{ -\frac{1}{2}\boldsymbol{b}_i^{B\top}(\boldsymbol{D}^{B^{-1}} + \boldsymbol{Z}_{i,(m_i)}^{B\top}\boldsymbol{W}_i\boldsymbol{Z}_{i,(m_i)}^B)\boldsymbol{b}_i^B + \boldsymbol{b}_i^{B\top}\boldsymbol{Z}_{i,(m_i)}^{B\top}\boldsymbol{W}_i(\boldsymbol{Y}_{i,(m_i)}^B - \boldsymbol{X}_{i,(m_i)}^B\boldsymbol{\beta}^B) + \delta_i\boldsymbol{b}_i^{B\top}\boldsymbol{\alpha}^B - e^{\boldsymbol{b}_i^{B\top}\boldsymbol{\alpha}^B}q_i\Big\} \quad (3.38)$$

where q_i is constant with respect to $\boldsymbol{b}_i^B$. To update the values of the random effects, we used the Metropolis-Hastings algorithm, taking the proposal density to be a multivariate normal density with mean obtained from few Newton-Raphson iterations applied to Equation (3.38) and covariance matrix equal to the inverse curvature evaluated at the mean of the proposal distribution.

3.5 Simulations

A simulation study was carried out to evaluate the performance of the proposed model, relatively to the performance of the LN-SREM(RE) and PH-SREM models, under specific MAR and MNAR dropout mechanisms. 500 datasets, each containing 1000 subjects, were simulated assuming a random intercept and slope model, with its parameters' values given in Subsection 3.2.4. The maximum study duration was set to 5 years, with the marker assessed every 3 months. The dropout time was assumed to follow the exponential distribution, depending on already observed values (the MAR scenario) or on both observed and unobserved measurements (the MNAR scenario). Under MAR, the hazard of dropout depended on the most recent observed marker value y_{ij}, i.e. $h(t) = \exp(3.956 - 0.221y_{ij})$, for $t_{ij} \leq t < t_{ij+1}$, whereas, under MNAR, the chance of dropout at a given time was, in addition, allowed to depend on future observed marker values y_{ij+1}, i.e. $h(t) = \exp\{3.956 - 0.221(y_{ij} + y_{ij+1})/2\}$, for $t_{ij} \leq t < t_{ij+1}$. Simulation of such dropout times is feasible using the Inverse CDF theorem. Marker's values measured after the simulated dropout time were ignored.

The examined models were: (1) the PH-SREM(CV) model described in Section 3.3, with $h(t|\boldsymbol{b};\boldsymbol{\theta}) = h_0(t;\boldsymbol{\psi})\exp\{\alpha m(t)\}$, where $\log\{h_0(t;\boldsymbol{\psi})\}$ is modeled through restricted cubic splines with 2 internal knots, (2) a SREM in which the hazard of dropout

depends explicitly on the random effects, $h(t|\boldsymbol{b};\boldsymbol{\theta}) = h_0(t;\boldsymbol{\psi})\exp(\alpha_0 b_{0i} + \alpha_1 b_{1i})$, referred to as the PH-SREM(RE) model, modeling $h_{0;\boldsymbol{\psi}}(t)$ as in PH-SREM(CV), (3) the LN-SREM(RE) model, (4) the proposed PH-SREM(LV,RE) model, $h\{t|\boldsymbol{b}, \mathcal{Y}_i(t);\boldsymbol{\theta}\} = h_0(t;\boldsymbol{\psi})\exp(\alpha_0' b_{0i} + \alpha_1' b_{1i} + \phi y_{ij})$, $t_{ij} \leq t < t_{ij+1}$, using a constant baseline hazard function within 4 intervals, and (5) an LMM. It should be noted that under the MNAR scenario, none of the examined models is correctly specified as dropout is assumed to depend on the mean of the current and future observed marker values, thus providing information on the robustness of the SREMs under misspecified MNAR dropout mechanisms. For both the PH-SREM(CV) and the PH-SREM(RE) model, the STJM software (Crowther, 2012) was used with 10 quadrature points.

In the proposed model, we used a burn-in period of 1000 iterations and recorded 15000 draws. To adjust for potential auto-correlation in the posterior samples, we thinned the chain by keeping every third draw, thus producing posterior inferences based on 5000 draws. The prior distributions used were the following:

- Fixed effects parameters: $\boldsymbol{\beta}^\top = (\beta_0, \beta_1)^{top} \sim N(\mathbf{0}, 100\boldsymbol{I}_2)$
- Survival submodel parameters: $(\alpha_0', \alpha_1', \phi)^{top} \sim N(\mathbf{0}, 100\boldsymbol{I}_3)$
- Baseline survival parameters: $h_k \sim \text{Gamma}(0.001, 0.001)$, $k = 1, \ldots, 4$
- Covariance matrix of the random effects: $\boldsymbol{D} \sim \text{InvWish}\{\text{diag}(20, 1), 2\}$
- Within-subjects precision: $\omega = 1/\sigma^2 \sim \text{Gamma}(0.001, 0.001)$

Results from the simulation study are presented in Table 3.1. In the MAR case, the LMM was nearly unbiased, as expected. All SREMs but the PH-SREM(LV,RE) model tended to overestimate the marker decline over time, with the bias being more severe (bias up to 86.59%) in the LN-SREM(RE) and PH-SREM(RE) models. The proposed model gave nearly unbiased results with the association parameters being on average close to zero and their 95% CI's including zero about 95% percent of the time. Under the MNAR scenario, the LMM estimates were biased, as expected. Focusing on the slope estimate, it was again almost unbiased in the proposed model, moderately

biased in PH-SREM(CV), whereas it was seriously biased in the remaining models. The Monte Carlo SEs were close to the corresponding average model-based SEs in all SREMs, though the LMM model-based SE was underestimated by 9.64% and 12.18% for the MAR and MNAR scenarios, respectively.

To examine the mixing properties of the MCMC algorithm used in the proposed model, we present trace plots of posterior samples of β_0, β_1 and $\omega = 1/\sigma^2$, and α_0, α_1 and ϕ in Figures 3.3 and 3.4, respectively. It can be seen that our algorithm mixes well and converges soon. There is some small degree of auto-correlation in the posterior samples (mainly for the population intercept and slope), which is reasonable considering the fact that the model is particularly complex. Recall, however, that the auto-correlation in the final sample have reduced since we have thinned the chain by keeping every third draw, thus producing posterior inferences based on 5000 draws.

To evaluate the performance of the proposed model for bivariate marker data relatively to a bivariate LMM, we ran an additional simulation study. 500 datasets, each containing 1000 subjects, were simulated including 2 markers, mimicking the CD4 and VL evolution. The maximum study was set to 5 years, with the markers assessed every 3 months. For the CD4 evolution we assumed a random intercept and slope model on the square root scale, $y^c_{ij} = (\beta^c_0 + b^c_{i0}) + (\beta^c_1 + b^c_{i1})t_{ij} + \epsilon^c_{ij}$, whereas for the viral load levels a model of the form $y^v_{ij} = (\beta^v_0 + b^v_{i0}) + (\beta^v_1 + b^v_{i1})t_{ij} + \beta^v_2 \log(t_{ij} + \kappa) + \epsilon^v_{ij}$ was used on the log10 scale. The true values of the population parameters were $\boldsymbol{\beta}^{c\top} = (23.712, -1.313)$ and $\boldsymbol{\beta}^{v\top} = (4.213, 0.190, -0.299)$. Note also that the term $\log(t_{ij} + \kappa)$ was used to capture the non-linear evolution of the viral load, i.e. VL tends to reach very high levels soon after infection, then it drops to some minimum levels after a few months and it increases again but at a much smaller rate. The value of κ was pre-specified through a grid search using results from linear mixed models fitted to CASCADE data, yielding $\kappa = 0.013$. The within-subject residuals, ϵ^c_{ij} and ϵ^v_{ij} , were assumed to follow independent normal distributions with zero means and variances σ^2_c and σ^2_v, respectively, whereas correlation between the two markers was induced through the correlation of the random effects, i.e. $\boldsymbol{b}^{B\top}_i = (b^c_{i0}, b^c_{i1}, b^v_{i0}, b^v_{i1})^{top} \sim N(\mathbf{0}, \boldsymbol{D}^B)$. The true value of the

Table 3.1: Performance of the different models under MAR and MNAR mechanisms: results from 500 replications with each dataset containing 1000 subjects. The mean estimate, percentage bias, empirical coverage probability and the Monte Carlo standard error are shown. The true values were 23.60 cells/μL and -1.30 cells/μL/year for the intercept and slope (on the square root scale), respectively. Estimates for the proposed PH-SREM(LV,RE) model are based on posterior modes.

Model	MAR				MNAR			
PH-SREM(CV)	Mean	Bias	Cov	MCSE	Mean	Bias	Cov	MCSE
Intercept	23.61	0.03	0.96	0.15	23.64	0.16	0.95	0.16
Slope	-1.65	27.06	0.00	0.08	-1.40	7.92	0.75	0.07
Association (α)	-0.25			0.01	-0.24			0.01
PH-SREM(RE)								
Intercept	23.68	0.33	0.96	0.14	23.71	0.46	0.88	0.16
Slope	-2.19	68.45	0.00	0.10	-1.88	44.61	0.00	0.09
Association intercept (α_0)	-0.26			0.01	-0.26			0.01
Association slope (α_1)	-0.73			0.06	-0.70			0.06
LN-SREM(RE)								
Intercept	23.69	0.36	0.94	0.15	23.71	0.45	0.90	0.16
Slope	-2.43	86.59	0.00	0.10	-2.09	60.46	0.00	0.10
Association intercept (λ_0)	0.16			0.01	0.17			0.01
Association slope (λ_1)	0.54			0.03	0.59			0.03
PH-SREM(LV,RE)								
Intercept	23.59	-0.04	0.96	0.15	23.65	0.19	0.96	0.16
Slope	-1.31	0.97	0.94	0.11	-1.29	-0.49	0.96	0.10
Association intercept (α_1')	-0.00		0.94	0.02	-0.07			0.02
Association slope (α_1')	-0.01		0.94	0.06	-0.17			0.06
Association observed CD4 (ϕ)	-0.22	-1.00	0.95	0.02	-0.15			0.02
Linear mixed model								
Intercept	23.60	-0.02	0.96	0.15	23.63	0.13	0.95	0.15
Slope	-1.30	0.32	0.92	0.07	-1.10	-15.19	0.15	0.07

Mean, Mean estimate; Cov, 95% Empirical coverage probability; Bias, % bias;

MCSE, Monte Carlo standard error

PH-SREM(CV): $h_i(t|\boldsymbol{b};\boldsymbol{\theta}) = h_0(t;\boldsymbol{\psi})\exp\left[\alpha\left\{(\beta_0+b_{0i})+(\beta_1+b_{1i})t\right\}\right],\ t>0$

PH-SREM(RE): $h_i(t|\boldsymbol{b};\boldsymbol{\theta}) = h_0(t;\boldsymbol{\psi})\exp\left(\alpha_0 b_{0i}+\alpha_1 b_{1i}\right),\ t>0$

LN-SREM(RE): $\log(T_i) = \zeta+\lambda_0 b_{i0}+\lambda_1 b_{i1}+r$

PH-SREM(LV,RE): $h_i\left\{t|\boldsymbol{b},\mathcal{Y}_i(t);\boldsymbol{\theta}\right\} = h_0(t;\boldsymbol{\psi})\exp\left(\alpha_0' b_{0i}+\alpha_1' b_{1i}+\phi y_{ij}\right),\ t_{ij}\le t<t_{ij+1}$

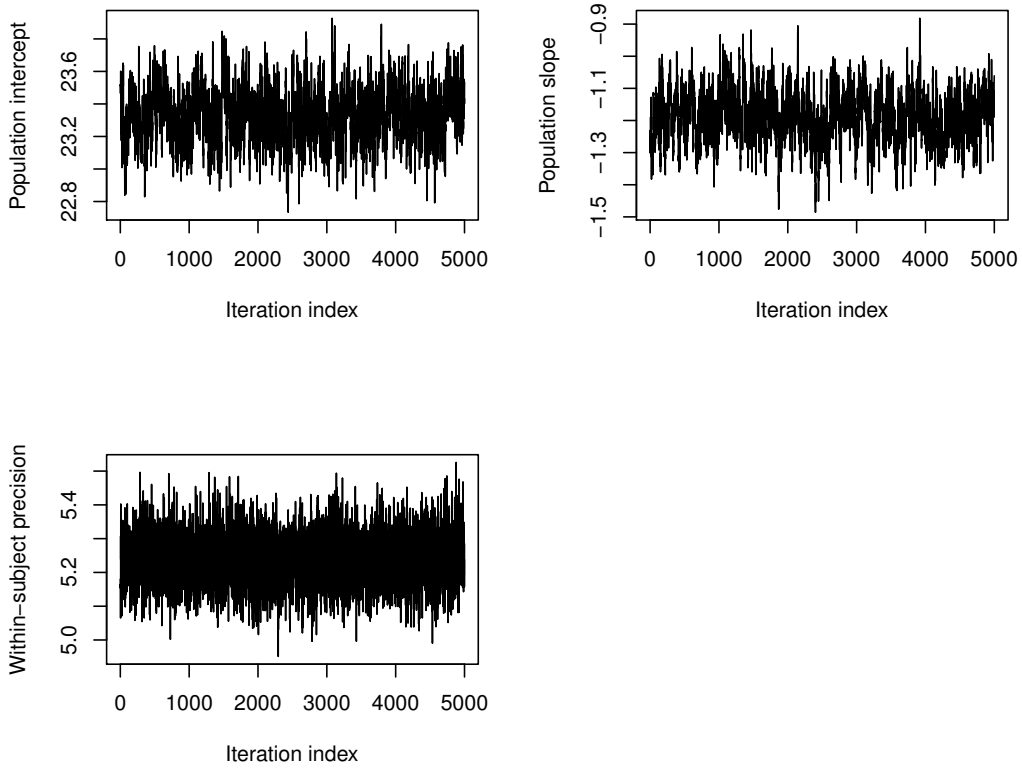


Figure 3.3: Trace plots of posterior samples for the population intercept, the population slope and the within-subject precision. The burn-in period includes 1000 iterations and the chain was thinned by keeping every third draw.

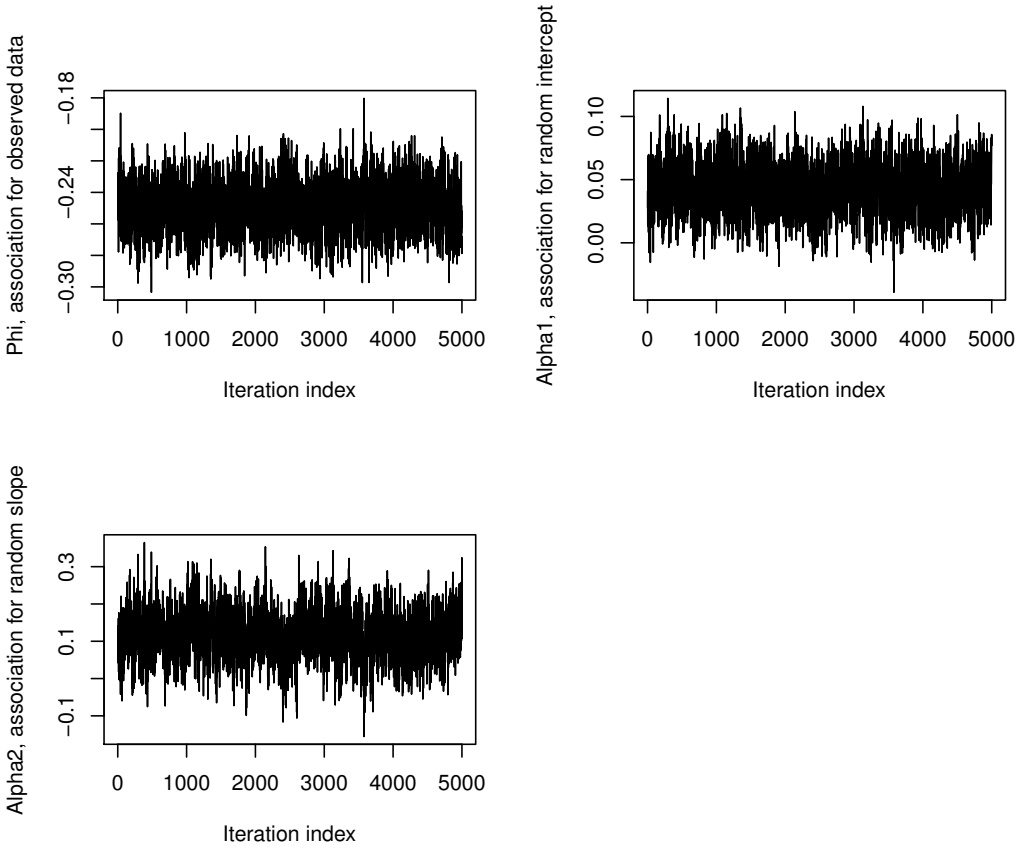


Figure 3.4: Trace plots of posterior samples for the parameters linking the hazard of dropout with the the observed data, the random intercept and the random slope, ϕ, α_0 and α_1, respectively. The burn-in period includes 1000 iterations and the chain was thinned by keeping every third draw.

covariance matrix of the random effects was

$$\boldsymbol{D}^B = \begin{pmatrix} 23.25 & -2.48 & -1.76 & 0.53 \\ -2.48 & 1.98 & 0.10 & -0.30 \\ -1.76 & 0.10 & 0.71 & -0.16 \\ 0.53 & -0.30 & -0.16 & 0.12 \end{pmatrix},$$

whereas the within-subjects variances for CD4 and VL were $\sigma_c^2 = 5.5$ and $\sigma_v^2 = 0.23$, respectively.

We considered both MAR and MNAR dropout mechanisms. Under the MAR scenario, the hazard of dropout was assumed to depend on the most recent markers' values, i.e. $h(t) = \exp\left(\lambda_1 + \phi_1^c y_{ij}^c + \phi_1^v y_{ij}^v\right)$ for $t_{ij} \leq t < t_{ij+1}$, whereas under the MNAR scenario, the hazard of dropout was also allowed to depend on future markers' values, i.e. $h(t) = \exp\left\{\lambda_1 + \phi_1^c(y_{ij}^c + y_{ij+1}^c)/2 + \phi_1^v(y_{ij}^v + y_{ij+1}^v)/2\right\}$ for $t_{ij} \leq t < t_{ij+1}$. Markers' values measured after the simulated dropout time were excluded.

We fitted the bivariate extension of the proposed model in each simulated dataset, and for comparison, a bivariate linear mixed model was also fitted. The marker model was correctly specified in both models. In the proposed model, the survival sub-model was assumed to be $h\left\{t|\mathcal{Y}_i^c(t), \mathcal{Y}_i^v(t), \boldsymbol{b}_i; \boldsymbol{\theta}\right\} = h_0(t; \boldsymbol{\psi}) \exp\left(\phi^c y_{ij}^c + \phi^v y_{ij}^v + \alpha_0^c b_{i0}^c + \alpha_1^c b_{i1}^c + +\alpha_1^v b_{i0}^v + \alpha_1^v b_{i1}^v\right)$ for $t_{ij} \leq t < t_{ij+1}$, with the baseline hazard function assumed to be constant within 4 intervals. The prior distributions used were the following:

- Fixed effects parameters: $\boldsymbol{\beta}^B \sim N(\mathbf{0}, 100\boldsymbol{I}_5)$
- Survival submodel parameters: $(\alpha_0^c, \alpha_1^c, \alpha_0^v, \alpha_1^v, \phi^c, \phi^v)^{top} \sim N(\mathbf{0}, 100\boldsymbol{I}_6)$
- Baseline survival parameters: $h_k \sim \text{Gamma}(0.001, 0.001)$, $k = 1, \ldots, 4$
- Covariance matrix of the random effects: $\boldsymbol{D} \sim \text{InvWish}\{\text{diag}(20, 1, 1, 1), 4\}$
- Within-subjects precision: $\omega^c = 1/\sigma_c^2 \sim \text{Gamma}(0.001, 0.001)$; $\omega^v = 1/\sigma_v^2 \sim \text{Gamma}(0.001, 0.001)$

Results from the simulation study are presented in Table 3.2. Under the MAR dropout mechanism, the bivariate linear mixed model was unbiased, as expected. The

proposed model had only small biases (up to 1.7%), while keeping the coverage rates close to the nominal levels. Under MNAR, the bivariate linear model produced biased results, which was also to be expected. The proposed model, though, yielded estimates with relatively small biases (up to 3.8%) again with coverage rates close to the nominal levels.

3.6 Application to the CASCADE data

We applied the previous models to markers' data from the CASCADE study, restricting to Europeans infected through sex between men after 1/1/2004. As we focus on markers' trends during untreated HIV infection, measurements taken after cART initiation or AIDS onset were excluded. Drop-out was assumed to have occurred on the first of the following dates: date of cART initiation, date of death or date of AIDS diagnosis, whereas study termination was regarded as a non-informative censoring event. As the most recent within the study period guidelines on cART initiation were based mainly on the current observed CD4 counts and partly on VL levels, and the vast majority of dropouts were due to cART initiation (98%), the MAR assumption seems plausible.

Initially, we fitted the models to CD4 data only. A random intercept and slope model on the square root scale was used for the marker model, with the time origin being the HIV seroconversion date. Age at seroconversion, in four groups ([15,25),[25,35),[35,45) and 45+), was used as a covariate in both the marker and dropout models. In the proposed model, we used linear splines with two knots to model the association between the hazard of dropout and the observed square root CD4 counts, whereas the baseline hazard was assumed to be constant within 20 intervals. We used the following prior distributions:

- Fixed effects parameters: $\boldsymbol{\beta}^\top = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)^{top} \sim N(\mathbf{0}, 100\boldsymbol{I}_5)$
- Survival submodel parameters: $(\gamma_1, \gamma_2, \gamma_3, \alpha_0', \alpha_1', \boldsymbol{\phi})^{top} \sim N(\mathbf{0}, 100\boldsymbol{I}_8)$
- Baseline survival parameters: $h_k \sim \text{Gamma}(0.001, 0.001)$, $k = 1, \ldots, 20$

Table 3.2: Performance of the bivariate proposed model and the bivariate LMM under MAR and MNAR mechanisms: results from 500 replications with each dataset containing 1000 subjects. The mean estimate, percentage bias, empirical coverage probability and the Monte Carlo standard error are shown. The true values of β_0^c and β_1^c are 23.712, -1.313, respectively, and the true values of β_0^v, β_1^v and β_2^v are 4.213, 0.190, -0.299, respectively. Estimates for the proposed PH-SREM(LV,RE) model are based on posterior modes.

Model	MAR				MNAR			
Proposed model	Mean	Bias	Cov	MCSE	Mean	Bias	Cov	MCSE
CD4: Intercept (β_0^c)	23.71	0.00	0.95	0.15	23.75	0.14	0.95	0.15
CD4: Slope (β_1^c)	-1.32	0.32	0.94	0.11	-1.29	-1.42	0.95	0.10
VL: Intercept (β_0^v)	4.21	-0.03	0.94	0.03	4.20	-0.19	0.93	0.03
VL: Slope-1 (β_1^v)	0.19	1.70	0.94	0.03	0.20	3.83	0.94	0.02
VL: Slope-2 (β_2^v)	-0.30	0.05	0.95	0.01	-0.30	0.42	0.95	0.01
CD4:Association intercept (α_0^c)	-0.00		0.95	0.02	-0.06			0.02
CD4: Association slope (α_1^c)	0.00		0.94	0.09	-0.13			0.08
VL:Association intercept (α_0^v)	0.02		0.94	0.12	0.11			0.11
VL: Association slope (α_1^v)	0.05		0.94	0.35	0.30			0.32
Association observed CD4 (ϕ_c)	-0.21	0.19	0.95	0.01	-0.15			0.01
Association observed VL (ϕ_v)	0.46	-1.50	0.96	0.06	0.36			0.06
Linear mixed model								
CD4: Intercept (β_0^c)	23.72	0.04	0.95	0.15	23.74	0.13	0.95	0.15
CD4: Slope (β_1^c)	-1.31	0.03	0.91	0.08	-1.13	-13.75	0.20	0.07
VL: Intercept (β_0^v)	4.21	0.00	0.93	0.03	4.21	0.04	0.93	0.03
VL: Slope-1 (β_1^v)	0.19	-0.47	0.92	0.02	0.16	-14.57	0.64	0.02
VL: Slope-2 (β_2^v)	-0.30	-0.07	0.94	0.01	-0.30	-0.39	0.96	0.01

Mean, Mean estimate; Cov, 95% Empirical coverage probability; Bias, (%) bias;

MCSE, Monte Carlo standard error

Proposed model: $h\left\{t|\mathcal{y}_i^c(t),\mathcal{y}_i^v(t),\boldsymbol{b}_i;\boldsymbol{\theta}\right\} = h_0(t;\boldsymbol{\psi})\exp\Big(\alpha_0^c b_{i0}^c + \alpha_1^c b_{i1}^c$
$+\alpha_0^v b_{i0}^v + \alpha_1^v b_{i1}^v + \phi^c y_{ij}^c + \phi^v y_{ij}^v\Big)$ for $t_{ij} \leq t < t_{ij+1}$.

- Covariance matrix of the random effects: $\boldsymbol{D} \sim \text{InvWish}\{\text{diag}(20, 1), 2\}$
- Within-subjects variance: $\omega \sim \text{Gamma}(0.001, 0.001)$

The first 2000 iterations were discarded as a burn-in period and we recorded 60000 draws thereafter. To reduce the auto-correlation in the posterior samples, we made inferences using 20000 iterations keeping every third draw. In both the PH-SREM(CV) and the PH-SREM(RE) models, we used 6 internal knots with 10 quadrature points.

The results are presented in Table 3.3. The effect of age was roughly similar across all models. However, LN-SREM(RE) and PH-SREM(RE) yielded much steeper CD4 decline estimates compared to the one derived from the LMM. The proposed model and the PH-SREM(CV) model gave slope estimates closer but still steeper than the one from the LMM.

As cART initiation depends also on VL, we extended the proposed model incorporating the VL levels (on the log10 scale) using natural cubic splines with 3 internal knots for the fixed effects and two random effects related to the baseline VL values and the time term, respectively. We used linear splines with one knot to relate dropout to the observed log10 VL levels. The following prior distributions were used:

- Fixed effects parameters: $\boldsymbol{\beta}^B \sim N(\mathbf{0}, 100\boldsymbol{I}_{13})$
- Survival submodel parameters: $(\gamma_1, \gamma_2, \gamma_3, \alpha_0^c, \alpha_1^c, \alpha_0^v, \alpha_1^v, \boldsymbol{\phi}^c, \boldsymbol{\phi}^v)^{top} \sim N(\mathbf{0}, 100\boldsymbol{I}_{12})$
- Baseline survival parameters: $h_k \sim \text{Gamma}(0.001, 0.001)$, $k = 1, \ldots, 20$
- Covariance matrix of the random effects: $\boldsymbol{D} \sim \text{InvWish}\{\text{diag}(20, 1, 1, 1), 4\}$
- Within-subjects precisions: $\omega^c = 1/\sigma_c^2 \sim \text{Gamma}(0.001, 0.001)$; $\omega^v = 1/\sigma_v^2 \sim \text{Gamma}(0.001, 0.001)$

A bivariate LMM was then fitted for comparison reasons. In the proposed model, the CD4 decline estimate (-1.39) is much closer to the corresponding one from the bivariate LMM (-1.27), with the parameters associating random CD4 effects with the hazard of dropout being very close to zero. Thus, under the assumed model, results suggest that

once the observed data of both markers have been taken into account, censoring of CD4 counts mainly due to cART initiation seems to be MAR. However, the corresponding association parameters for VL levels are far from zero, implying that the probability of dropout may not be fully explained by the observed CD4 and VL measurements; it might also depend on the underlying subject-specific VL trajectories but not on the underlying CD4 slope. Note that, these conclusions are conditional on the model being correctly specified and thus additional sensitivity analyses on key aspects of the model are required to back up the conclusions.

3.7 Discussion

In this chapter we investigated the performance of a specific class of joint models, termed shared random effects models, in estimating the rate of change of longitudinal markers subject to MAR dropout. Recall that as described in Chapter 2, SREMs are joint models in which the marker and the time-to-dropout process are assumed to be independent given the random effects, generally corresponding to an MNAR assumption about the drop-out process. We analytically calculated the bias in the population slope estimate by both the LN-SREM(RE) and PH-SREM(CV) models, using specific true MAR dropout mechanisms based on our motivating example. The results showed that the asymptotic bias in the estimated slope increases as the MAR dropout mechanism strengthens. In estimating asymptotic bias, the examined SREM models assumed continuous dropout times, whereas the data were generated under discrete dropout times. One could argue that the bias in estimated slope was, at least partly, due to this model misspecification. However, further bias calculations after resolving this misspecification showed that this is unlikely; what is more, our analytical results were further supported by our simulation study findings. Motivated by the definition of MNAR, we proposed an alternative model that allows the hazard of dropout to depend on the last observed marker value, as well as on the random effects. Due to computational difficulties, we did not run similar bias calculations for the proposed model. However, we evaluated

Table 3.3: Modeling temporal trends in CD4 cell counts and VL levels in the CASCADE data: Results from various SREMs and linear mixed models. Estimates for the proposed PH-SREM(LV,RE) models are based on posterior modes. All parameters apart from the association ones refer to the longitudinal marker model.

Model	Estimate	(SE)	95%CI
LMM			
Age [15,25) Ref.	23.88	0.15	(23.58,24.18)
Age [25,35)	-0.06	0.17	(-0.40,0.28)
Age [35,45)	-0.45	0.18	(-0.81,-0.09)
Age [45,)	-0.86	0.22	(-1.29,-0.42)
Slope	-1.26	0.03	(-1.32,-1.20)
PH-SREM(LV,RE) (Univariate)			
Age [15,25) Ref.	24.14	0.16	(23.83,24.45)
Age [25,35)	-0.01	0.18	(-0.36,0.34)
Age [35,45)	-0.40	0.18	(-0.75,-0.04)
Age [45,)	-0.83	0.23	(-1.28,-0.38)
Slope	-1.64	0.04	(-1.73,-1.56)
Assoc. intercept (α_0')	-0.01	0.01	(-0.03,0.01)
Assoc. slope (α_1')	-0.36	0.03	(-0.42,-0.29)
Assoc. obs. CD4 (ϕ_1)	-0.11	0.01	(-0.14,-0.09)
Assoc. obs. CD4 (ϕ_2)	-0.33	0.01	(-0.35,-0.31)
Assoc. obs. CD4 (ϕ_3)	-0.05	0.01	(-0.07,-0.03)
PH-SREM(CV)			
Age [15,25) Ref.	24.00	0.16	(23.70,24.31)
Age [25,35)	0.01	0.18	(-0.36,0.34)
Age [35,45)	-0.42	0.19	(-0.78,-0.05)
Age [45,)	-0.87	0.23	(-1.32,-0.42)
Slope	-1.58	0.04	(-1.65,-1.51)
Assoc. (α)	-0.27	0.01	(-0.28,-0.26)
PH-SREM(RE)			
Age [15,25) Ref.	24.28	0.16	(23.98,24.59)
Age [25,35)	0.03	0.18	(-0.32,0.38)
Age [35,45)	-0.34	0.19	(-0.71,0.03)
Age [45,)	-0.79	0.23	(-1.24,-0.34)
Slope	-2.08	0.04	(-2.15,-2.00)
Assoc. intercept (α_0)	-0.25	0.01	(-0.26,-0.24)
Assoc. slope (α_1)	-0.95	0.03	(-1.01,-0.89)
LN-SREM(RE)			
Age [15,25) Ref.	24.39	0.16	(24.08,24.70)
Age [25,35)	0.07	0.18	(-0.29,0.43)
Age [35,45)	-0.29	0.19	(-0.67,0.09)
Age [45,)	-0.73	0.24	(-1.19,-0.27)
Slope	-2.59	0.05	(-2.68,-2.50)
Assoc. intercept (λ_0)	0.10	0.00	(0.10,0.11)
Assoc. slope (λ_1)	0.50	0.01	(0.48,0.52)

Model	Estimate	(SE)	95%CI
Bivariate LMM			
Age [15,25) Ref.	24.14	0.14	(23.86,24.42)
Age [25,35)	-0.33	0.16	(-0.64,-0.02)
Age [35,45)	-0.73	0.17	(-1.06,-0.41)
Age [45,)	-1.19	0.21	(-1.60,-0.77)
Slope (CD4)	-1.27	0.03	(-1.33,-1.21)
PH-SREM(LV,RE) (Bivariate)			
Age [15,25)	24.21	0.14	(23.93,24.49)
Age [25,35)	-0.30	0.16	(-0.63,0.01)
Age [35,45)	-0.71	0.17	(-1.04,-0.39)
Age [45,)	-1.19	0.22	(-1.62,-0.76)
Slope (CD4)	-1.39	0.05	(-1.48,-1.30)
Assoc. intercept (α_0^c)	0.05	0.01	(0.03,0.07)
Assoc. slope (α_1^c)	0.05	0.04	(-0.04,0.13)
Assoc. intercept (α_0^v)	0.77	0.06	(0.65,0.88)
Assoc. slope (α_1^v)	2.17	0.21	(1.76,2.58)
Assoc. obs. CD4 (ϕ_1^c)	-0.12	0.01	(-0.14,-0.09)
Assoc. obs. CD4 (ϕ_2^c)	-0.35	0.01	(-0.37,-0.33)
Assoc. obs. CD4 (ϕ_3^c)	-0.07	0.01	(-0.09,-0.04)
Assoc. obs. VL (ϕ_1^v)	-0.79	0.04	(-0.87,-0.71)
Assoc. obs. VL (ϕ_2^v)	0.25	0.03	(0.19,0.32)

PH-SREM(LV,RE): $h_i(t) = h_0(t)\exp\left\{\alpha_0' b_{0i} + \alpha_1' b_{1i} + \phi_1 y_{ij} I(y_{ij} \leq \sqrt{250}) + \phi_2 y_{ij} I(\sqrt{250} < y_{ij} \leq \sqrt{550}) + \phi_3 y_{ij} I(y_{ij} > \sqrt{550})\right\}$

PH-SREM(CV): $h_i(t) = h_0(t)\exp\left[\alpha\left\{(\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t\right\}\right],\ t > 0$

PH-SREM(RE): $h_i(t) = h_0(t)\exp\left(\alpha_0 b_{0i} + \alpha_1 b_{1i}\right),$

LN-SREM(RE): $\log(T_i) = \zeta + \lambda_0 b_{i0} + \lambda_1 b_{i1} + r$

Biv. PH-SREM(LV,RE): $h_i(t) = h_0(t)\exp\left\{\alpha_0^c b_{0i}^c + \alpha_1^c b_{1i}^c + \phi_1^c y_{ij}^c I(y_{ij}^c \leq \sqrt{250}) + \phi_2^c y_{ij}^c I(\sqrt{250} < y_{ij}^c \leq \sqrt{550}) + \phi_3^c y_{ij}^c I(y_{ij}^c > \sqrt{550}) + \alpha_0^v b_{0i}^v + \alpha_1^v b_{1i}^v + \phi_1^v y_{ij}^v I(y_{ij}^v \leq 3.5) + \phi_2^v y_{ij}^v I(y_{ij}^v > 3.5)\right\}$, for $t_{ij} \leq t < t_{ij+1}$.

Superscripts c and v stand for CD4 and VL, respectively. VL evolution is modeled through natural splines.

its comparative performance in simulation studies. Using specific MAR and MNAR dropout scenarios, the proposed model had negligible bias in the estimated slope under both scenarios; the PH-SREM(CV) model yielded a slope estimate with serious bias under MAR and moderate bias under MNAR. To the contrary, the SREMs that use a random-effects parameterization yielded seriously biased results in all cases. It should be noted, though, that the robustness of the proposed model might be partly attributed to the type of the assumed true MNAR dropout mechanism. That is, $(y_{ij} + y_{ij+1})/2$ could be decomposed in y_{ij} and $(y_{ij+1} - y_{ij})/2$, with the second part being probably not that far from the random slope as the evolution of the marker is assumed linear. Nevertheless, in the MNAR bivariate simulations, the proposed model yielded VL longitudinal parameter estimates with small biases at most (up to 3.83%), although the true evolution of VL was highly nonlinear.

We fitted the examined models to recent data from the CASCADE study, with cART being the dominating reason for dropout. All SREMs, including the proposed one, yielded a steeper CD4 decline compared to the estimate from the LMM, with the SREMs using a random-effects parameterization yielding the most extreme results. It should be highlighted, though, that in the proposed dropout model, the observed CD4 are modeled through linear splines, whereas the PH-SREM(CV) model assumes log-linear relationship between the hazard of dropout and the underlying CD4 values. Thus, differences between the results of these two models might be partly due to the different functional form used for modeling CD4. As cART initiation depends not only on CD4 counts but also on VL levels, a bivariate version of the proposed model was applied and its results were compared to those from a corresponding bivariate LMM. The two bivariate models yielded similar CD4 slope estimates, indicating that, after considering also VL levels, the dropout probabilities are unlikely to depend on the underlying CD4 slope.

There are certain extensions that could be incorporated in the proposed methodology. First, subjects can drop out of the study for various other reasons (e.g. due to AIDS onset or death) and this should have been taken into account. Competing-risk

joint models, similar to the one proposed by Huang et al. (2010), could be used to account for that. However, in our motivating example, this would increase the computational complexity unnecessarily, as the proportion of subjects developing AIDS or dying prior to cART initiation is practically negligible. In this work we only investigated the performance of specific SREMs under specific MAR dropout mechanisms. Thus, results may not hold for other joint models or other types of MAR dropout. In addition, other types of model misspecification have not been extensively examined here.

To summarize, although SREMs have been effectively used to obtain unbiased estimates under MNAR dropout, we showed that they can lead to seriously biased slope estimates under specific forms of MAR dropout; SREMs using a random-effect parameterization had the worst performance and should thus be avoided, at least in settings similar to ours. When SREM(CV) models are applied in cases where the MAR assumption seems reasonable, results should be interpreted with caution. Our proposed model, which performed well under specific MAR and MNAR dropout mechanisms, could be applied alternatively, at least as part of a sensitivity analysis, as discrimination between MAR and MNAR is not possible.

3. PERFORMANCE OF LONGITUDINAL AND TIME-TO-DROPOUT JOINT MODELS WHEN THE DROPOUT MECHANISM IS AT RANDOM

Chapter 4

Misspecifying the covariance structure in a linear mixed model under MAR dropout

4.1 Introduction

In the previous chapter we examined the case where SREM models are fitted to incomplete MAR data. We analytically showed that the bias in the estimated slope can be substantial and it is related to the amount of MAR dropout. To overcome this difficulty, we proposed an alternative SREM model that is more robust under specific both MAR and MNAR dropout mechanisms. However, another source of bias in the estimates of the population average evolution is due to misspecification of the covariance structure. That is, even when missingness is MAR, likelihood-based modeling of the observed marker data can provide valid estimates only under the condition that the whole likelihood model is correct (Rubin, 1976), implying that both the mean evolution over time and the covariance structure are correctly specified.

In our motivating example of modeling the CD4 count evolution during the HIV natural history, CD4 measurements taken after ART initiation or AIDS onset are by

definition excluded, leading to a monotone missing data problem in the form of dropout. Treatment guidelines up to 2015 suggested ART initiation when CD4 counts drop below a certain level, thus, missing CD4 data are likely to be MAR, as the dropout probabilities depend on the observed CD4 counts. Under the MAR assumption, modeling the observed CD4 data is sufficient to obtain valid estimates. However, previous application of LMMs to real CD4 data (Stirrup et al., 2015) has shown that the choice of the covariance structure greatly influences the estimation of the CD4 average rate of change over time (slope).

LMMs that include random intercept and slope have been typically used by many researchers to model pre-ART CD4 count data. However, to improve model fit, it has been suggested to either use splines in the design matrix of the random effects (Rizopoulos, 2012b) or to use a random intercept and slope model along with an additional stochastic process such as Brownian motion (BM) (Stirrup et al., 2015, Taylor and Law, 1998). In the missing data literature, other alternative methods which relax the distributional assumptions about the covariance structure and/or other aspects of the model have been proposed (Seaman and Copas, 2009, Tsiatis, 2007), including doubly robust generalized estimating equations for longitudinal data (Seaman and Copas, 2009); for an overview of non/semi-parametric methods under missing data, see (Tsiatis, 2007). However, even under MAR missingness, these methods require explicit modeling of the missingness mechanism to compute the inverse probability weights. In our motivating example, and most probably in other disciplines, (i) the dropout times are usually continuous, (ii) there are multiple dropout mechanisms of potentially different nature (MCAR, MAR, or MNAR), and (iii) the visiting times are highly irregular resulting in strongly imbalanced data. Thus, the latter methods cannot be routinely applied in CD4 natural history data. Other methods assuming a Normal distribution with unstructured covariance have also been proposed (Lu and Mehrotra, 2010), but they cannot be easily applied to HIV cohort data due to highly irregular clinic visits. Given these difficulties, one can focus on LMMs with/without additional stochastic processes, which can be quite flexible and do not require modeling of potentially complex dropout

mechanisms.

When comparing the fit of nested LMMs, likelihood ratio tests are frequently applied. In cases where the null hypothesis lies in the boundary of the parameter space (e.g. when testing multiple random effects or any subset of them), likelihood ratio tests should be modified by using, for example, mixtures of chi-square distributions (Self and Liang, 1987) or permutation tests (Drikvandi et al., 2012). However, when comparing models that are not nested (e.g. an LMM with splines to an LMM using random intercept and slope along with a BM process), likelihood ratio tests are ruled out. Other information criteria such as the AIC or the BIC criterion, can be also used, but they rely on asymptotics without a straightforward probabilistic interpretation. An additional difficulty arises from the fact that it is not easy to define the penalty term in the BIC criterion for mixed models (Keselman et al., 1998, Spiegelhalter et al., 2002b). Alternatively, one can use Bayesian model comparison, an approach that uses the marginal likelihoods of the compared models to derive the evidence in favor of one model against another through their respective posterior model probabilities (PMPs). Analytic evaluation of the integrals involved in the computation of the marginal likelihoods is almost always impossible. To overcome this difficulty, several procedures based on Markov chain Monte Carlo methods have been proposed (Han and Carlin, 2001), however these approaches can be very time-consuming and usually require a substantial amount of programming effort. Methods to numerically approximate PMPs have been previously proposed for LMMs (Saville and Herring, 2009), but they are based on Laplace approximations without analytic integration over the fixed effects. Also, to our knowledge, extensions to LMMs with additional stochastic processes (e.g. the BM process) have not been implemented so far.

It is known that the PMPs can be sensitive to the choice of prior distributions (Kass and Raftery, 1995), which is particularly important when there is no prior information on the parameters to be estimated. Default priors with large variances could be used, though it should be kept in mind that, as the prior variance increases, the PMP tends to favor the null model (Kass and Raftery, 1995).

In this chapter, a) we analytically show that using a covariance structure that is simpler than the true one can lead to seriously biased population parameters estimates of an LMM under MAR dropout and b) we adopt a fully Bayesian model comparison approach based on the posterior model probabilities to discriminate between alternative models. In Section 4.2, we analytically derive the bias in the marker rate of change over time (slope) estimate, when the covariance structure is missspecified, as a function of the model parameters for one specific MAR dropout mechanism. We also compare, under misspecified covariance structure, the induced bias in three modeling approaches: (i) an LMM with random intercept and slope; (ii) an LMM with natural splines in the design matrix of the random effects; and (iii) an LMM with random intercept and slope along with a stochastic fractional BM process. In Section 4.4, we extend the previous setting by quantifying the bias in the estimated slope difference between two groups. In Section 4.5, we describe the proposed Bayesian model comparison approach, and we evaluate its performance relative to other frequently used approaches in a simulation study. All examined models are also fitted to recent data from the CASCADE study. We compare their performance using all considered criteria. Section 4.8 presents concluding remarks along with a discussion of limitations and possible extensions.

4.2 Asymptotic Bias in the Population Average Marker Rate of Change Estimate Due to Covariance Structure Misspecification under MAR Drop-out

In this section, we assume the same data generating mechanism as that described in Sections 3.2 and 3.2.2. That is, $\boldsymbol{Y} = (Y_1, \ldots, Y_Q)^\top$ denote the marker measurements intended to be collected at fixed scheduled times $t_1, \ldots, t_Q$, with M denoting the number of the observed marker measurements, modelled through a logistic MAR dropout mechanism of the form $P(M = j | M \geq j, \boldsymbol{Y}_{(j)}; c_1, c_2) = \frac{e^{c_1 + c_2(Y_j - Y^\star)}}{1 + e^{c_1 + c_2(Y_j - Y^\star)}}$, for $j < Q$, and $P(M = Q | M \geq Q, \boldsymbol{Y}_{(Q)}; c_1, c_2) = 1$.

Suppose that the "true average" marker evolution over time can be described

by a multivariate Normal model with mean $E_0(\boldsymbol{Y}) = \boldsymbol{X}\boldsymbol{\beta}_0$ and covariance matrix $Var_0(\boldsymbol{Y}) = \sigma_0^2\boldsymbol{V}_0$. Subscript "0" indicates the true parameter values and the whole parameter vector of the true model is denoted by $\boldsymbol{\theta}_{L0}^\top = (\boldsymbol{\beta}_0^\top, \boldsymbol{\theta}_{v0}^\top)$, where $\boldsymbol{\theta}_{v0}$ denotes the true values of the variance components (i.e. σ_0^2 along with the parameters in $\boldsymbol{V}_0$). Let us assume that a multivariate Normal model is fitted to the data using the correct design matrix $\boldsymbol{X}$, i.e. $E(\boldsymbol{Y}) = \boldsymbol{X}\boldsymbol{\beta}$, but using a wrong covariance structure, $Var(\boldsymbol{Y}) = \sigma^2\boldsymbol{V}$, with $\boldsymbol{\theta}_L^\top = (\boldsymbol{\beta}^\top, \boldsymbol{\theta}_v^\top)$ denoting the parameter vector of the misspecified model. We study the asymptotic bias of the maximum likelihood estimator under the misspecified model, $\widehat{\boldsymbol{\theta}_L}$, by first deriving the limit in probability of the estimator $\widehat{\boldsymbol{\theta}_L}$. Using standard asymptotic theory (Newey and McFadden, 1994), under mild regularity conditions, the estimator $\widehat{\boldsymbol{\theta}_L}$ converges to the value of $\boldsymbol{\theta}_L$ solving the equation $E_0\left\{\frac{\partial}{\partial\boldsymbol{\theta}_L}\log f(\boldsymbol{Y}_{(M)};\boldsymbol{\theta}_L)\right\} = \boldsymbol{0}$, with the expectation over the true joint distribution of the dropout mechanism and the observed data $(M, \boldsymbol{Y}_{(M)})$, evaluated at the true parameter values [i.e. $\boldsymbol{\theta}_0^\top = (\boldsymbol{\beta}_0^\top, \boldsymbol{\theta}_{v0}^\top, c_1, c_2)$, with $\boldsymbol{\theta}$ denoting the full parameter vector of the joint distribution of the marker and dropout processes]. Note that the score vector of the misspecified model for a random individual from the population is equal to $\frac{\partial}{\partial\boldsymbol{\theta}_L}\sum_{j=1}^{Q} I(M=j)\left\{-\log|\sigma^2\boldsymbol{V}_{(j)}|/2 - (\boldsymbol{Y}_{(j)} - \boldsymbol{X}_{(j)}\boldsymbol{\beta})^\top\boldsymbol{V}_{(j)}^{-1}(\boldsymbol{Y}_{(j)} - \boldsymbol{X}_{(j)}\boldsymbol{\beta})/2\sigma^2\right\}$, where $\boldsymbol{X}_{(j)}$ and $\boldsymbol{V}_{(j)}$ are the appropriate submatrices of $\boldsymbol{X}$ and $\boldsymbol{V}$ for the jth dropout pattern, respectively. Our approach is similar in spirit with that of Saha and Jones (2005), who calculated the asymptotic bias in the parameters of an LMM under MNAR missingness. Since $\log f(\boldsymbol{Y}_{(M)};\boldsymbol{\theta}_L)$ is a quadratic form in $\boldsymbol{Y}_{(j)}$, it can be shown that the system of equations $E_0\left\{\frac{\partial}{\partial\boldsymbol{\theta}_L}\log f(\boldsymbol{Y}_{(M)};\boldsymbol{\theta}_L)\right\} = \boldsymbol{0}$ simplifies to

$$\frac{\partial}{\partial\boldsymbol{\theta}_L}\sum_{j=1}^{Q}\left[-\frac{a_{0j}\log|\sigma^2\boldsymbol{V}_{(j)}|}{2} - \frac{a_{0j}}{2\sigma^2}\left\{\mathrm{tr}(\boldsymbol{V}_{(j)}^{-1}\boldsymbol{\Omega}_{0(j)}) + (\boldsymbol{\mu}_{0(j)} - \boldsymbol{X}_{(j)}\boldsymbol{\beta})^\top\boldsymbol{V}_{(j)}^{-1}(\boldsymbol{\mu}_{0(j)} - \boldsymbol{X}_{(j)}\boldsymbol{\beta})\right\}\right] = \boldsymbol{0}, \tag{4.1}$$

with $\boldsymbol{\mu}_{0(j)}$ and $\boldsymbol{\Omega}_{0(j)}$ being the true mean and covariance matrix of $\boldsymbol{Y}_{(j)}|M=j;\boldsymbol{\theta}_0$, respectively, and $a_{0j} = P(M=j;\boldsymbol{\theta}_0)$ the true marginal probability of dropout of the jth dropout pattern. $\boldsymbol{\mu}_{0(j)}$, $\boldsymbol{\Omega}_{0(j)}$ and $a_{0j} = P(M=j;\boldsymbol{\theta}_0)$ are not available in closed

form unless missing data are MCAR; to estimate them, one could use quasi-Monte Carlo integration coupled with importance sampling (see Section 3.2.3 of Chapter 3). We show through direct calculations that, for given $\boldsymbol{\theta}_v$, the bias in the population parameters' estimates from the misspecified model equals

$$\left(\sum_{j=1}^{Q} a_{0j}\boldsymbol{X}_{(j)}^{\top}\boldsymbol{V}_{(j)}^{-1}\boldsymbol{X}_{(j)}\right)^{-1}\sum_{j=1}^{Q} a_{0j}\boldsymbol{X}_{(j)}^{\top}(\boldsymbol{V}_{(j)}^{-1}-\boldsymbol{V}_{0(j)}^{-1})(\boldsymbol{\mu}_{0(j)}-\boldsymbol{X}_{(j)}\boldsymbol{\beta}_0).$$

Thus, the magnitude of the bias depends on both (i) the difference between the assumed and "true" covariance matrix $\boldsymbol{V}_{(j)}-\boldsymbol{V}_{0(j)}$ and (ii) the difference between $\boldsymbol{\mu}_{0(j)}$ and $\boldsymbol{X}_{(j)}\boldsymbol{\beta}_0$. This implies that if the assumed covariance matrix $\boldsymbol{V}_{(j)}$ is not equal to the "true" one $\boldsymbol{V}_{0(j)}$, the bias is equal to zero only if the dropout mechanism is MCAR (i.e. $\boldsymbol{\mu}_{0(j)}=\boldsymbol{X}_{(j)}\boldsymbol{\beta}_0$, as $\boldsymbol{Y}$ and M are independent under MCAR).

To prove this result, we differentiate Equation (4.1) over $\boldsymbol{\beta}$, and we obtain

$$\sum_{j=1}^{Q} a_{0j}\boldsymbol{X}_{(j)}^{\top}\boldsymbol{V}_{(j)}^{-1}\boldsymbol{X}_{(j)}\boldsymbol{\beta}-\sum_{j=1}^{Q} a_{0j}\boldsymbol{X}_{(j)}^{\top}\boldsymbol{V}_{(j)}^{-1}\boldsymbol{\mu}_{0(j)}=\mathbf{0}, \tag{4.2}$$

which has an explicit solution with respect to $\boldsymbol{\beta}$. We now use the fundamental result from the missing data theory stating that the estimates from the correct likelihood model are consistent under any MAR dropout mechanism, implying that the expectation of the score vector from the correct model over $(M,\boldsymbol{Y}_{(M)})$ is zero at the true parameter values. This, in turn, implies that

$$\sum_{j=1}^{Q} a_{0j}\boldsymbol{X}_{(j)}^{\top}\boldsymbol{V}_{0(j)}^{-1}\boldsymbol{X}_{(j)}\boldsymbol{\beta}_0-\sum_{j=1}^{Q} a_{0j}\boldsymbol{X}_{(j)}^{\top}\boldsymbol{V}_{0(j)}^{-1}\boldsymbol{\mu}_{0(j)}=\mathbf{0},$$

where $\boldsymbol{V}_{0(j)}$ corresponds to the covariance matrix for the jth dropout pattern under the correct model. Writing $\boldsymbol{\mu}_{0(j)}$ as $\boldsymbol{X}_{(j)}\boldsymbol{\beta}_0+\boldsymbol{\Delta}_{(j)}$, we show that

$$\begin{aligned}\sum_{j=1}^{Q} a_{0j}\boldsymbol{X}_{(j)}^{\top}\boldsymbol{V}_{0(j)}^{-1}\boldsymbol{X}_{(j)}\boldsymbol{\beta}_0 &= \sum_{j=1}^{Q} a_{0j}\boldsymbol{X}_{(j)}^{\top}\boldsymbol{V}_{0(j)}^{-1}\boldsymbol{X}_{(j)}\boldsymbol{\beta}_0+\sum_{j=1}^{Q} a_{0j}\boldsymbol{X}_{(j)}^{\top}\boldsymbol{V}_{0(j)}^{-1}\boldsymbol{\Delta}_{(j)}\\ \Rightarrow \sum_{j=1}^{Q} a_{0j}\boldsymbol{X}_{(j)}^{\top}\boldsymbol{V}_{0(j)}^{-1}\boldsymbol{\Delta}_{(j)} &= \mathbf{0}.\end{aligned}$$

Therefore, based on Equation (4.2), it follows that

$$\begin{aligned}\sum_{j=1}^{Q} a_{0j}\boldsymbol{X}_{(j)}^{\top}\boldsymbol{V}_{(j)}^{-1}\boldsymbol{X}_{(j)}(\boldsymbol{\beta}-\boldsymbol{\beta}_0) &= \sum_{j=1}^{Q} a_{0j}\boldsymbol{X}_{(j)}^{\top}\boldsymbol{V}_{(j)}^{-1}\boldsymbol{\Delta}_{(j)} \\ &= \sum_{j=1}^{Q} a_{0j}\boldsymbol{X}_{(j)}^{\top}(\boldsymbol{V}_{0(j)}^{-1}+\boldsymbol{V}_{(j)}^{-1}-\boldsymbol{V}_{0(j)}^{-1})\boldsymbol{\Delta}_{(j)},\end{aligned}$$

which, as $\sum_{j=1}^{Q} a_{0j}\boldsymbol{X}_{(j)}^{\top}\boldsymbol{V}_{0(j)}^{-1}\boldsymbol{\Delta}_{(j)}$ is always zero, implies that the bias in the estimated population parameters is equal to

$$\boldsymbol{\beta}-\boldsymbol{\beta}_0 = \left(\sum_{j=1}^{Q} a_{0j}\boldsymbol{X}_{(j)}^{\top}\boldsymbol{V}_{(j)}^{-1}\boldsymbol{X}_{(j)}\right)^{-1}\sum_{j=1}^{Q} a_{0j}\boldsymbol{X}_{(j)}^{\top}(\boldsymbol{V}_{(j)}^{-1}-\boldsymbol{V}_{0(j)}^{-1})(\boldsymbol{\mu}_{0(j)}-\boldsymbol{X}_{(j)}\boldsymbol{\beta}_0). \quad (4.3)$$

However, to properly derive the bias, Equation (4.1) should be solved with respect to the whole parameter vector, $\boldsymbol{\theta}_L$, e.g. through a Newton-Raphson algorithm. Results at convergence of the algorithm constitute the values to which $\widehat{\boldsymbol{\theta}_L}$ asymptotically converges. By comparing these values with the true parameter values, we calculate the bias in the estimates of the misspecified model.

4.2.1 Numerical Evaluation of the Bias in the Population Average Marker Rate of Change Estimate

As a specific application of Equation (4.1), we numerically evaluated the bias in the population slope estimate under misspecified covariance structure in three specific models:

- Model 1 (M1): LMM including a fractional BM process on top of random intercept and slope: $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta}_1 + \boldsymbol{Z}_1\boldsymbol{b}_1 + \boldsymbol{W} + \boldsymbol{\epsilon}_1$, where $\boldsymbol{b}_1 \sim N(\boldsymbol{0}, \sigma_1^2\boldsymbol{D}_1)$ denotes the random effects (i.e. random intercept and slope) and $\boldsymbol{\epsilon}_1 \sim N(\boldsymbol{0}, \sigma_1^2\boldsymbol{I}_{(Q)})$ the within-subject residuals, with $\boldsymbol{X}$ and $\boldsymbol{Z}_1$ being the design matrices of the fixed and random effects under M1, respectively. $\boldsymbol{W}$ denotes a fractional BM process defined at times $t_1, \ldots, t_Q$ with mean zero and covariance matrix $\sigma_1^2\boldsymbol{\Sigma}$, parametrized in terms of two scalar parameters κ and H (Stirrup et al., 2015); $H = 0.5$ implies a non-fractional BM process.

- Model 2 (M2): LMM with natural splines for the random effects: $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta}_2 + \boldsymbol{Z}_2\boldsymbol{b}_2 + \boldsymbol{\epsilon}_2$, where $\boldsymbol{b}_2 \sim N(\boldsymbol{0}, \sigma_2^2\boldsymbol{D}_2)$ and $\boldsymbol{\epsilon}_2 \sim N(\boldsymbol{0}, \sigma_2^2\boldsymbol{I}_{(Q)})$. $\boldsymbol{Z}_2$ denotes a natural spline basis matrix with one internal knot at 1.42 years and boundary knots at 0 and the maximum follow-up time, following the approach described in (Lambert and Royston, 2009).

- Model 3 (M3): LMM with random intercept and slope: $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta}_3 + \boldsymbol{Z}_3\boldsymbol{b}_3 + \boldsymbol{\epsilon}_3$, where $\boldsymbol{b}_3 \sim N(\boldsymbol{0}, \sigma_3^2\boldsymbol{D}_3)$ and $\boldsymbol{\epsilon}_3 \sim N(\boldsymbol{0}, \sigma_3^2\boldsymbol{I}_{(Q)})$. $\boldsymbol{Z}_3$ includes a column of ones along with a column with the measurement times (i.e. it is equal to $\boldsymbol{Z}_1$).

Two scenarios regarding the true marker model were investigated:

1. Assuming M1 as the true model, the models M2 and M3 were to be fitted. The true parameter values of M1 were $\boldsymbol{\beta}_{01} = (23.24, -1.12)^\top$, $\sigma_{01}^2 = 3.66$, $\kappa_0 = 6.82$, $H_0 = 0.31$, and $Var_0(b_{11}) = 23.66$, $Var_0(b_{12}) = 0.75$ and $Cov_0(b_{11}, b_{12}) = -2.02$, where $\boldsymbol{b}_1 = (b_{11}, b_{12})^\top$ denotes the random effects of M1. These values were chosen to mimic the CD4 count evolution on square root scale, based on results from models applied to the CASCADE data. In M2, we used one internal knot placed at 1.42 years (i.e. the median observation time in CASCADE) so as the number of covariance parameters in M1 and M2 to be comparable (5 and 6, respectively).

2. Assuming M2 as the true model, the models M1 and M3 were to be fitted. The true parameter values of M2 were $\boldsymbol{\beta}_{02} = (23.24, -1.12)^\top$, $\sigma_{02}^2 = 5.06$ and $Var_0(b_{21}) = 23.51$, $Var_0(b_{22}) = 2.90$, $Var_0(b_{23}) = 0.72$, $Cov_0(b_{21}, b_{22}) = 0.98$, $Cov_0(b_{21}, b_{23}) = 0.97$, $Cov_0(b_{22}, b_{23}) = 0.55$, where $\boldsymbol{b}_2 = (b_{21}, b_{22}, b_{23})^\top$ denotes the random effects of M2.

The maximum study duration was assumed to be 5 years, with the marker assessed every 4 months. To quantify the bias, we solved Equation (4.1) using various values for the dropout parameters. We set c_1 to correspond to the hazard of dropout at 500 CD4 cells/μL, which approximately corresponds to the median CD4 counts at sero-conversion (i.e. the baseline). We used 3 scenarios regarding the baseline hazard of

dropout (1%, 10% and 20%) and 21 points for c_1 $(-0.40, -0.38, \ldots, 0)$, leading to 63 scenarios. When M1 was the true model generating the data, over the 3 baseline hazard scenarios, the median (IQR, min-max) expected dropout by the end of the study was 29.2% (18.9%-41.1%, 14.0%-50.8%), 84.0% (83.1%-84.3%, 79.4%-84.6%), and 94.3% (92.6%-96.2%, 91.4%-96.8%) for 1%, 10%, and 20% baseline hazard of dropout, respectively. The corresponding median (IQR, min-max) expected numbers of the observed measurements were 13.9 (12.9-14.6, 11.8-14.9), 7.4 (7.1-7.8, 6.9-8.1), and 5.1 (4.9-5.2, 4.8-5.3), respectively. The corresponding dropout rates under M2 were identical to the first decimal digit. The results regarding the percentage bias in the estimated slope [i.e. $100 \times$ (Estimated $-$ True)/True] are presented in Figure 4.1. When $c_2 = 0$, which corresponds to an MCAR dropout mechanism, the bias in all cases was exactly equal to zero, confirming the theoretical findings. For a given value of c_2, the bias was smallest when the baseline dropout rate was lowest (1% at baseline), and increased as the baseline dropout rate increased. For a given baseline dropout rate, the bias increased with increasing absolute values of c_2 (i.e. increasing distance from 0, indicating heavier dropout). An exception occurred when M1 was the true model and the baseline dropout rate was high (20% at baseline), in which case the bias in the estimated slope by M3 tended to decrease when c_2 became lower than -0.20 (panel A_1 of Figure 4.1). The highest biases were observed when the fitted model had covariance structure simpler than the true model's (i.e. fitting an LMM with random intercept and slope when the true model was an LMM with a fractional BM process on top of random intercept and slope or an LMM with splines for the random effects; panels A_1 and B_1 of Figure 4.1, respectively). The bias was more pronounced when dropout was intense (bias up to 15%). Instead, the models with more complex covariance structures (i.e. models M1 and M2), performed roughly similarly well, with M1 slightly outperforming M2 (panels B_2 and A_2 of Figure 4.1; bias up to 6%).

We also considered the case where the maximum number of observations (Q) is doubled and halved using the same parameters for the dropout mechanism and the same maximum study duration. In general, most findings were similar; the highest

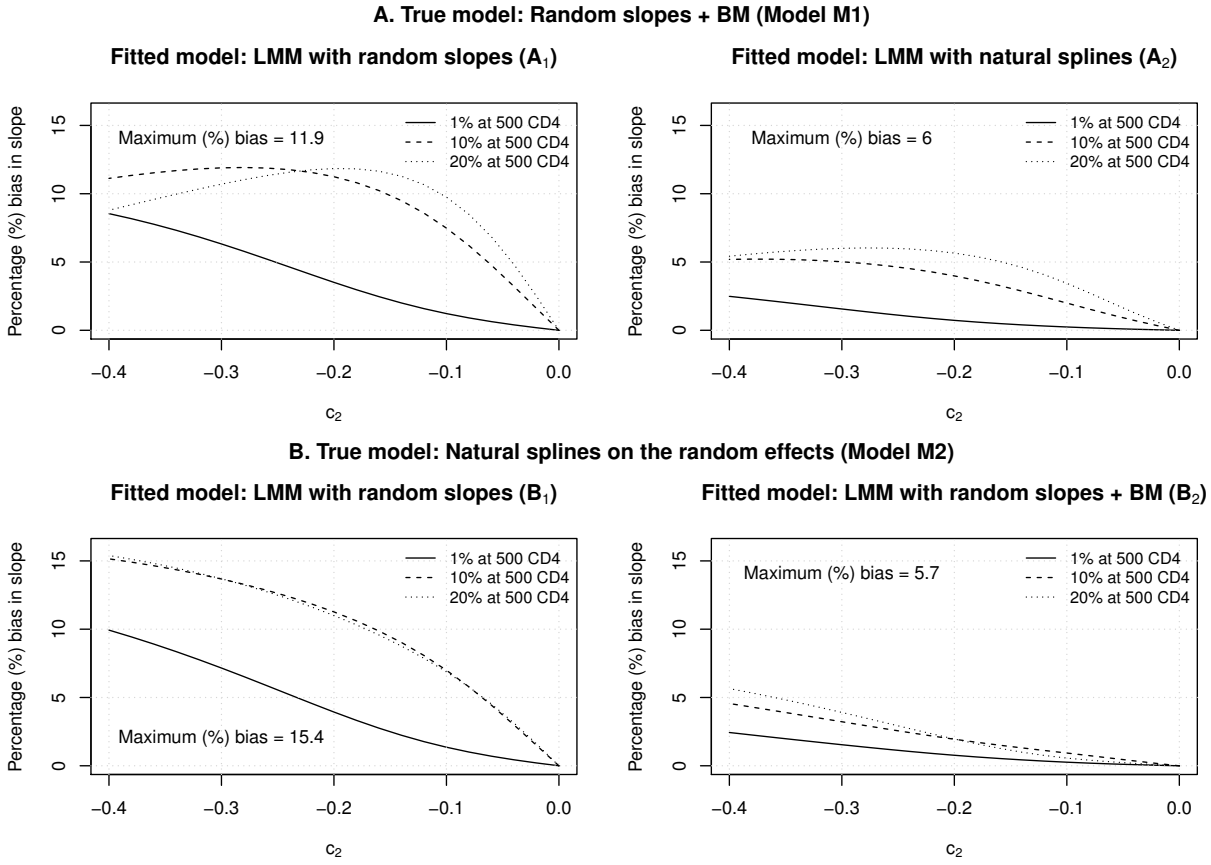


Figure 4.1: Asymptotic bias in a marker rate of change (slope) estimate under misspecified covariance structure. **A**. True covariance structure based on an LMM with a fractional BM process on top of random intercepts and slopes (Model M1), while the fitted models are an LMM with random intercept and slope (Model M3), panel A_1, and an LMM with natural splines for the random effects (Model M2), panel A_2. **B**. True covariance structure described by M2 and fitted covariance structure by M3 (B_1) and M1 (B_2). The parameter c_2 measures the change in the log-odds of dropout associated with one unit decrease in the current marker value. Three scenarios regarding the dropout rate at 500 CD4/μL, i.e. approximately the CD4 counts at seroconversion (baseline), were presented: 1%, 10% and 20%.

biases were seen in the estimates from M3 while the trend of increasing bias at higher dropout rates was also similar. However, the magnitude of the bias tended to increase when the maximum number of measurements was doubled ($Q = 32$ measurements) and decrease when it was halved ($Q = 8$ measurements). Specifically, for M3, the maximum percentage bias in the estimated slope increased to 19.3% when $Q = 32$ and decreased to 10.7% when $Q = 8$. Further investigation of the differences in the dropout rates between the scenarios of $Q = 8$, 16, and 32, under M1, revealed that the median (IQR, min-max) percentage of missing observations out of the ones intended to be collected tended to significantly increase at Q = 32, 69.2% (31.0%-78.8%, 14.1%-84.4%), and to decrease at Q = 8, 35.8% (11.7%- 49.2%, 3.4%-52.6%), compared to the case of $Q =$ 16, 53.5% (20.3%-67.3%, 7.2%-69.7%). Also, the variance estimates (evaluated at the scheduled times) from M3 were more seriously biased when $Q = 32$ as compared to when $Q = 8$. Since the degree of bias in the population parameters depends on the differences between the true variance matrix $\boldsymbol{V}_{0(j)}$ and the assumed one $\boldsymbol{V}_{(j)}$ for each dropout pattern, we may expect the bias to be higher at $Q = 32$ (where the summation involves more terms) if the assumed covariance structure poorly describes the true covariance in the data. However, the effect of the frequency of visits on the bias of the estimated slope requires more thorough investigation to draw a solid conclusion.

So far, we have assumed that the "true average" marker evolution followed a simple linear trend over time. To investigate the extent of bias due to covariance structure misspecification under a more complex mean marker evolution, we examined two additional settings:

1. we assumed that the "true average" marker evolution was piece-wise linear with one knot at 1.42 years, with the initial slope equal to -1.12 and the ultimate (i.e. time $>$ 1.42 years) slope equal to -1.22. The fixed effects slope was assumed to change at the same time point in the fitted models. In general, conclusions were similar to those previously reported. The detailed results are presented in Figures 4.2-4.3. It should be mentioned, though, that the highest biases were observed for the estimated ultimate slope, as dropout was heavier at later times.

For the ultimate slope, the model M3 had the highest biases (up to around 60%) yielding substantially steeper rate of decline compared to the true one, whereas M1 performed better that M2 (biases up to 3% and 20%, respectively). Regarding the initial slope estimates, the model M3 had again the highest bias, yielding negatively biased estimates (bias up to -17%).

2. we assumed that the true mean marker evolution was described by natural cubic splines. Full results are provided in Figure 4.4; most conclusions were in line with our previous findings.

4.3 Asymptotic bias in the population-averaged marker trend estimates due to covariance structure misspecification under continuous time MAR drop-out

The extent of bias was also evaluated under a continuous-time dropout mechanism, i.e. allowing the hazard of dropout to take place at any point in continuous time. Let T be the dropout time and $\delta = I(T < t_Q)$ be the dropout indicator, which means that $T \in [t_j, t_{j+1})$ implies that $M = j$. We assume that the hazard of dropout at time $t \in [t_j, t_{j+1})$ depends on the last observed marker value through the model $h(t|y_j, \eta_1, \eta_2) = \exp(\eta_1 + \eta_2 y_j)$. To find the limit in probability of the estimator $\hat{\boldsymbol{\theta}}_L$ of the misspecified model, we need to calculate

$$\sum_{j=1}^{Q} \int_{t_j}^{t_{j+1}} \int \frac{\partial}{\partial \boldsymbol{\theta}_L} \log f(\boldsymbol{Y}_{(j)}; \boldsymbol{\theta}_L) f_0(\boldsymbol{Y}_{(j)}, t; \boldsymbol{\theta}_{L0}, \eta_1, \eta_2) d\boldsymbol{Y}_{(j)} dt, \tag{4.4}$$

where $f(\boldsymbol{Y}_{(j)}; \boldsymbol{\theta}_L)$ denotes the likelihood of the misspecified model, whereas $f_0(\boldsymbol{Y}_{(j)}, t; \boldsymbol{\theta}_{L0}, \eta_1, \eta_2)$ denotes the likelihood of the "true" joint distribution of the marker and the dropout process, evaluated at the true parameter values. Interchanging the order of differentiation and integration, we show that Equation (4.4) is equal to

$$\frac{\partial}{\partial \boldsymbol{\theta}_L} \sum_{j=1}^{Q} \int \log f(\boldsymbol{Y}_{(j)}; \boldsymbol{\theta}_L) f_0(\boldsymbol{Y}_{(j)}, t_j \leq T < t_{j+1}; \boldsymbol{\theta}_{L0}, \eta_1, \eta_2) d\boldsymbol{Y}_{(j)}, \tag{4.5}$$

4.3 Asymptotic bias in the population-averaged marker trend estimates due to covariance structure misspecification under continuous time MAR drop-out

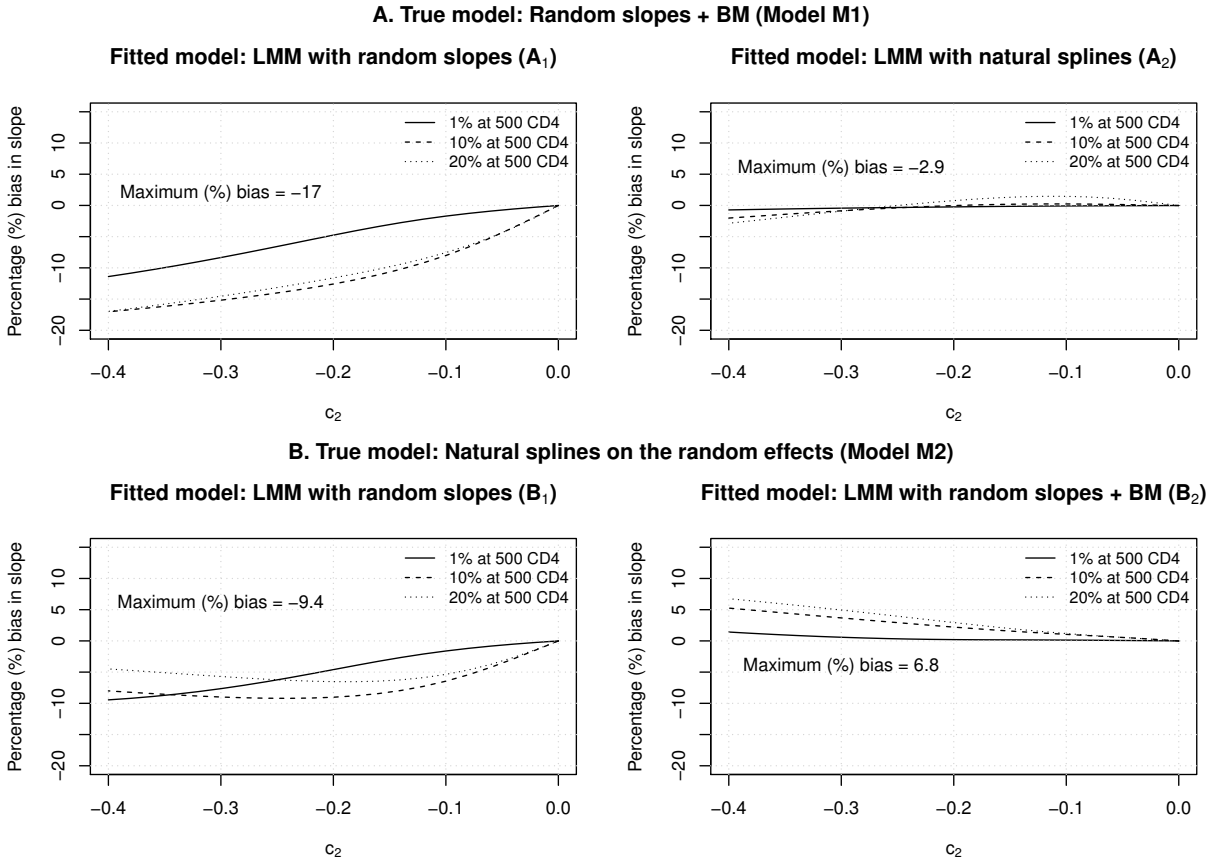


Figure 4.2: Asymptotic bias in the estimate of the initial slope of a marker (i.e. the rate of change up to 1.42 years since baseline) under misspecified covariance structure. **A**. True covariance structure based on an LMM with a fractional BM process on top of random intercepts and slopes (Model M1), while the fitted models are an LMM with random intercept and slope (Model M3), panel A_1, and an LMM with natural splines for the random effects (Model M2), panel A_2. **B**. True covariance structure described by M2 and fitted covariance structure by M3 (B_1) and M1 (B_2). The parameter c_2 measures the change in the log-odds of dropout associated with one unit decrease in the current marker value. Three scenarios regarding the dropout rate at 500 CD4/μL, i.e. approximately the CD4 counts at seroconversion (baseline), were presented: 1%, 10% and 20%.

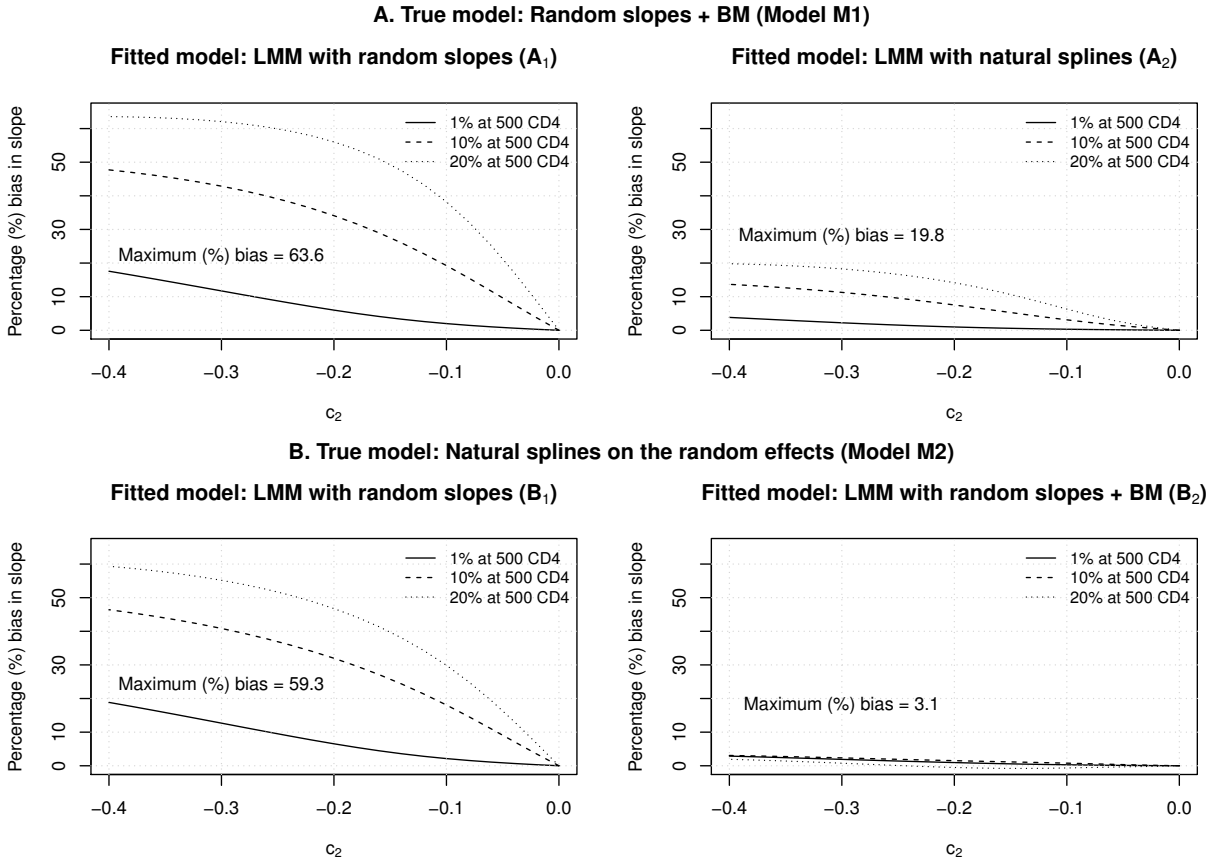


Figure 4.3: Asymptotic bias in the estimate of the ultimate slope of a marker (i.e. the rate of change after 1.42 years since baseline) under misspecified covariance structure. **A**. True covariance structure based on an LMM with a fractional BM process on top of random intercepts and slopes (Model M1), while the fitted models are an LMM with random intercept and slope (Model M3), panel A_1, and an LMM with natural splines for the random effects (Model M2), panel A_2. **B**. True covariance structure described by M2 and fitted covariance structure by M3 (B_1) and M1 (B_2). The parameter c_2 measures the change in the log-odds of dropout associated with one unit decrease in the current marker value. Three scenarios regarding the dropout rate at 500 CD4/μL, i.e. approximately the CD4 counts at seroconversion (baseline), were presented: 1%, 10% and 20%.

4.3 Asymptotic bias in the population-averaged marker trend estimates due to covariance structure misspecification under continuous time MAR drop-out

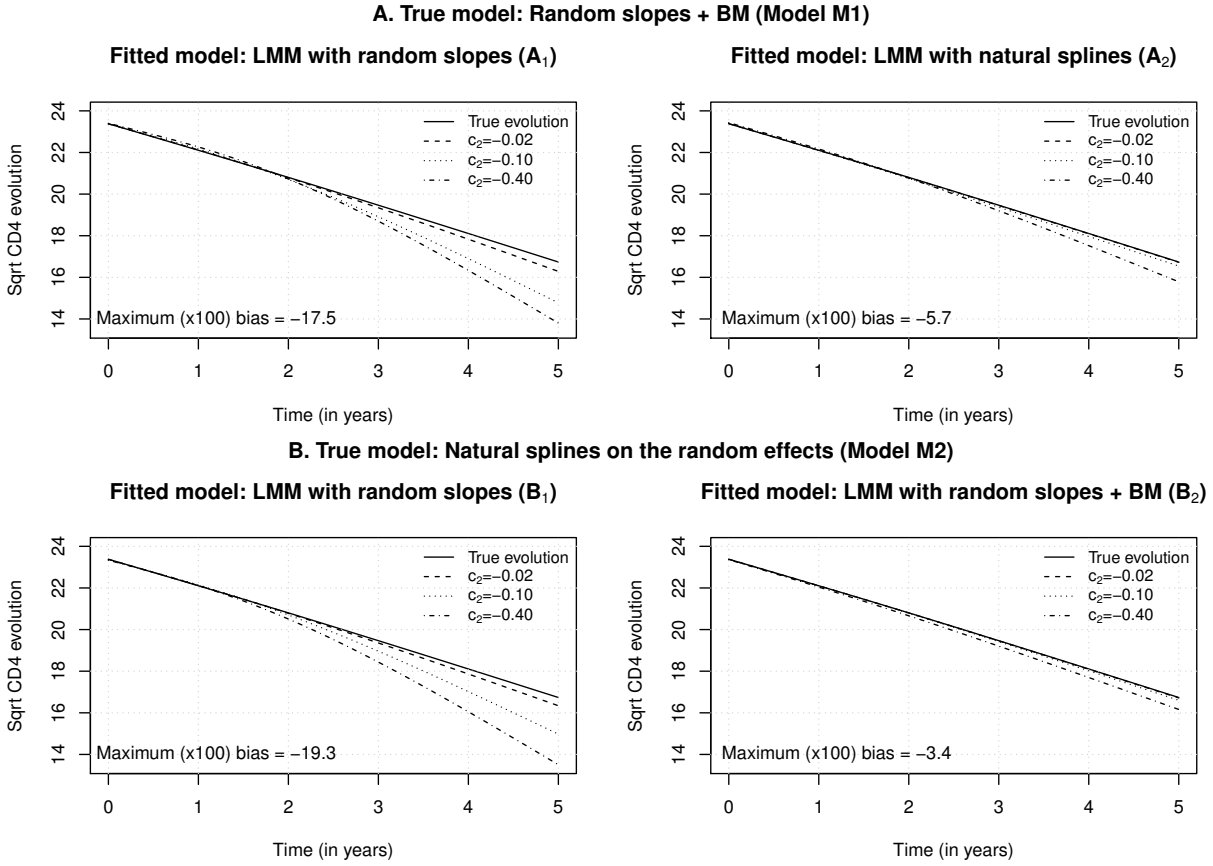


Figure 4.4: Asymptotic bias in a marker average evolution estimate under misspecified covariance structure. **A**. True covariance structure based on an LMM with a fractional BM process on top of random intercepts and slopes (Model M1), while the fitted models are an LMM with random intercept and slope (Model M3), panel A_1, and an LMM with natural splines for the random effects (Model M2), panel A_2. **B**. True covariance structure described by M2 and fitted covariance structure by M3 (B_1) and M1 (B_2). The solid curve represents the "true" marker evolution, whereas the dashed, dotted and dot-dashed curves show the estimated marker evolution for c_2 equal to -0.02, -0.10, -0.40, respectively. The parameter c_2 measures the change in the log-odds of dropout associated with one unit decrease in the current marker value. The dropout rate at 500 CD4/μL, i.e. approximately the CD4 counts at seroconversion (baseline), was 20%.

where $f_0(\boldsymbol{Y}_{(j)}, t_j \leq T < t_{j+1}; \boldsymbol{\theta}_{L0}, \eta_1, \eta_2)$ is a shorthand for $\int_{t_j}^{t_{j+1}} f_0(\boldsymbol{Y}_{(j)}, t; \boldsymbol{\theta}_{L0}, \eta_1, \eta_2)dt$. Noting that $f_0(\boldsymbol{Y}_{(j)}, t_j \leq T < t_{j+1}; \boldsymbol{\theta}_{L0}, \eta_1, \eta_2)$ can be factorized as $f_0(\boldsymbol{Y}_{(j)}|t_j \leq T < t_{j+1}; \boldsymbol{\theta}_{L0}, \eta_1, \eta_2)P(t_j \leq T < t_{j+1}; \boldsymbol{\theta}_{L0}, \eta_1, \eta_2)$ and letting $a_{0j} = P(t_j \leq T < t_{j+1}; \boldsymbol{\theta}_{L0}, \eta_1, \eta_2)$, it is evident that Equation (4.5) is equal to

$$\frac{\partial}{\partial \boldsymbol{\theta}_L} \sum_{j=1}^{Q} a_{0j} E_{\boldsymbol{Y}_{(j)}|t_j \leq T < t_{j+1}; \boldsymbol{\theta}_{L0}, \eta_1, \eta_2} \left\{ \log f(\boldsymbol{Y}_{(j)}; \boldsymbol{\theta}_L) \right\}, \tag{4.6}$$

which is, in turn, equal to

$$\frac{\partial}{\partial \boldsymbol{\theta}_L} \sum_{j=1}^{Q} \left[-\frac{a_{0j} \log |\sigma^2 \boldsymbol{V}_{(j)}|}{2} - \frac{a_{0j}}{2\sigma^2} \left\{ \text{tr}(\boldsymbol{V}_{(j)}^{-1} \boldsymbol{\Omega}_{0(j)}) + (\boldsymbol{\mu}_{0(j)} - \boldsymbol{X}_{(j)}\boldsymbol{\beta})^\top \boldsymbol{V}_{(j)}^{-1} (\boldsymbol{\mu}_{0(j)} - \boldsymbol{X}_{(j)}\boldsymbol{\beta}) \right\} \right], \tag{4.7}$$

where $\boldsymbol{\mu}_{0(j)} = E_0(\boldsymbol{Y}_{(j)}|t_j \leq T < t_{j+1}; \boldsymbol{\theta}_{L0}, \eta_1, \eta_2)$ and $\boldsymbol{\Omega}_{0(j)} = Var_0(\boldsymbol{Y}_{(j)}|t_j \leq T < t_{j+1}; \boldsymbol{\theta}_{L0}, \eta_1, \eta_2)$. These expectations cannot be obtained in closed form; we approximated them through importance sampling coupled with quasi-Monte Carlo integration following the procedure described in subsection 3.2.3. Note also that $P(t_j \leq T < t_{j+1}|\boldsymbol{Y}_{(j)}; \eta_1, \eta_2) = \exp\left\{ -\sum_{k=1}^{j-1} e^{\eta_1 + \eta_2 Y_k}(t_{k+1} - t_k) \right\} - \exp\left\{ -\sum_{k=1}^{j} e^{\eta_1 + \eta_2 Y_k}(t_{k+1} - t_k) \right\}$ and $P(T \geq t_Q|\boldsymbol{Y}_{(j)}, \eta_1, \eta_2) = \exp\left\{ -\sum_{k=1}^{Q-1} e^{\eta_1 + \eta_2 Y_k}(t_{k+1} - t_k) \right\}$.

The results are presented in Figures 4.5-4.8. The magnitude of the biases remained very similar to that obtained using the discrete logistic dropout model.

4.4 Asymptotic Bias in the Slope Difference Estimator Due to Covariance Structure Misspecification under MAR Dropout

In this section, we extend the previous setting by considering a two group scenario, with $p_g = \Pr(G = g)$, $g = 0, 1$, being the probability of being in group g ($p_0 + p_1 = 1$). For simplicity, we allow for group-specific marker slopes, but we assume that both groups have the same population average baseline value. Thus, the fixed-effects design matrix, $\boldsymbol{X}_{(j),g}$, now equals $\begin{pmatrix} \mathbf{1} & \boldsymbol{t}_{(j)} & I(g=1)\boldsymbol{t}_{(j)} \end{pmatrix}$, where $\boldsymbol{t}_{(j)}^\top = (t_1, \ldots, t_j)$ are the measurement times up to visit j. The corresponding population parameter vector, $\boldsymbol{\beta}_G$,

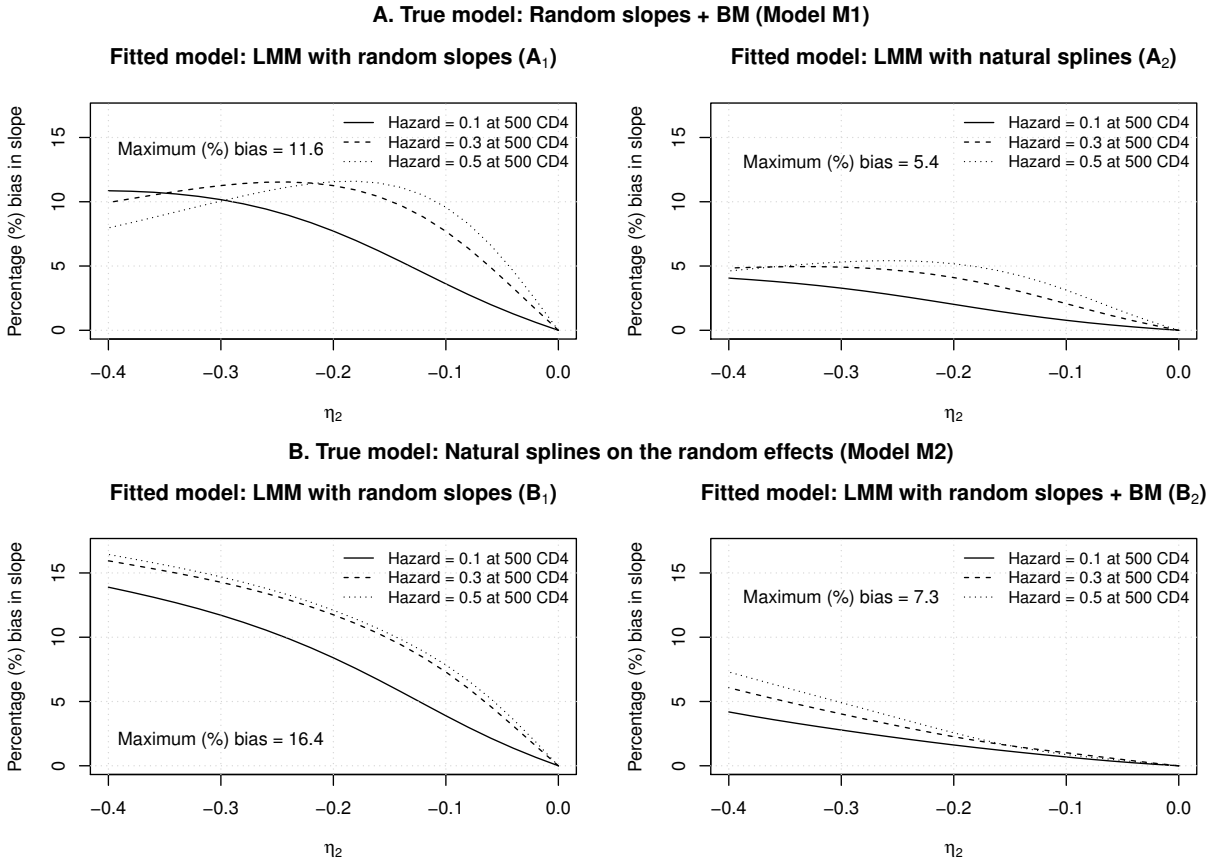


Figure 4.5: Asymptotic bias in a marker rate of change (slope) estimate under misspecified covariance structure and continuous-time dropout. **A**. True covariance structure based on an LMM with a fractional BM process on top of random intercepts and slopes (Model M1), while the fitted models are an LMM with random intercept and slope (Model M3), panel A_1, and an LMM with natural splines for the random effects (Model M2), panel A_2. **B**. True covariance structure described by M2 and fitted covariance structure by M3 (B_1) and M1 (B_2). The parameter η_2 measures the change in the log hazard rate of dropout associated with one unit decrease in the current marker value. Three scenarios regarding the hazard dropout rate at 500 CD4/μL, i.e. approximately the CD4 counts at seroconversion (baseline), were presented: 0.1, 0.30 and 0.50.

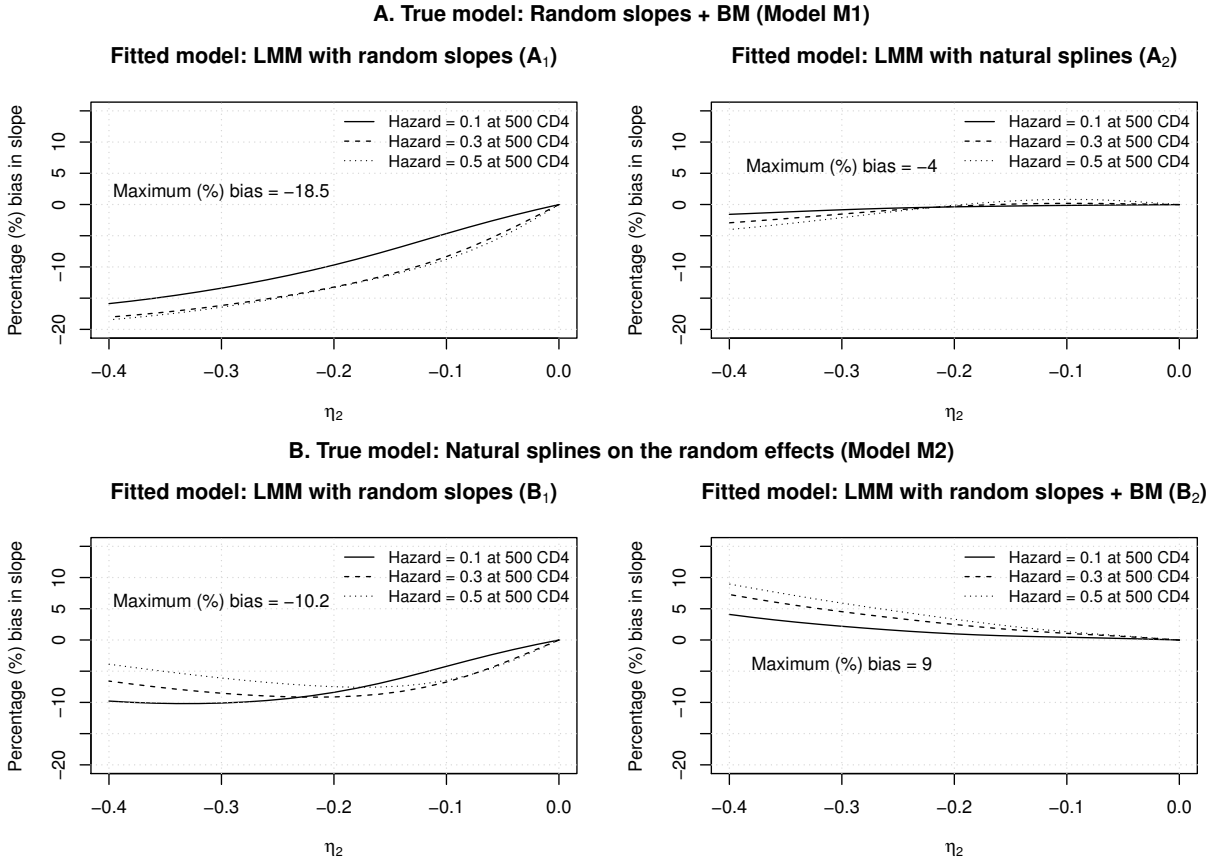


Figure 4.6: Asymptotic bias in the estimate of the initial slope of a marker (i.e. the rate of change up to 1.42 years since baseline) under misspecified covariance structure and continuous-time dropout. **A**. True covariance structure based on an LMM with a fractional BM process on top of random intercepts and slopes (Model M1), while the fitted models are an LMM with random intercept and slope (Model M3), panel A_1, and an LMM with natural splines for the random effects (Model M2), panel A_2. **B**. True covariance structure described by M2 and fitted covariance structure by M3 (B_1) and M1 (B_2). The parameter η_2 measures the change in the log hazard rate of dropout associated with one unit decrease in the current marker value. Three scenarios regarding the hazard dropout rate at 500 CD4/μL, i.e. approximately the CD4 counts at seroconversion (baseline), were presented: 0.1, 0.30 and 0.50.

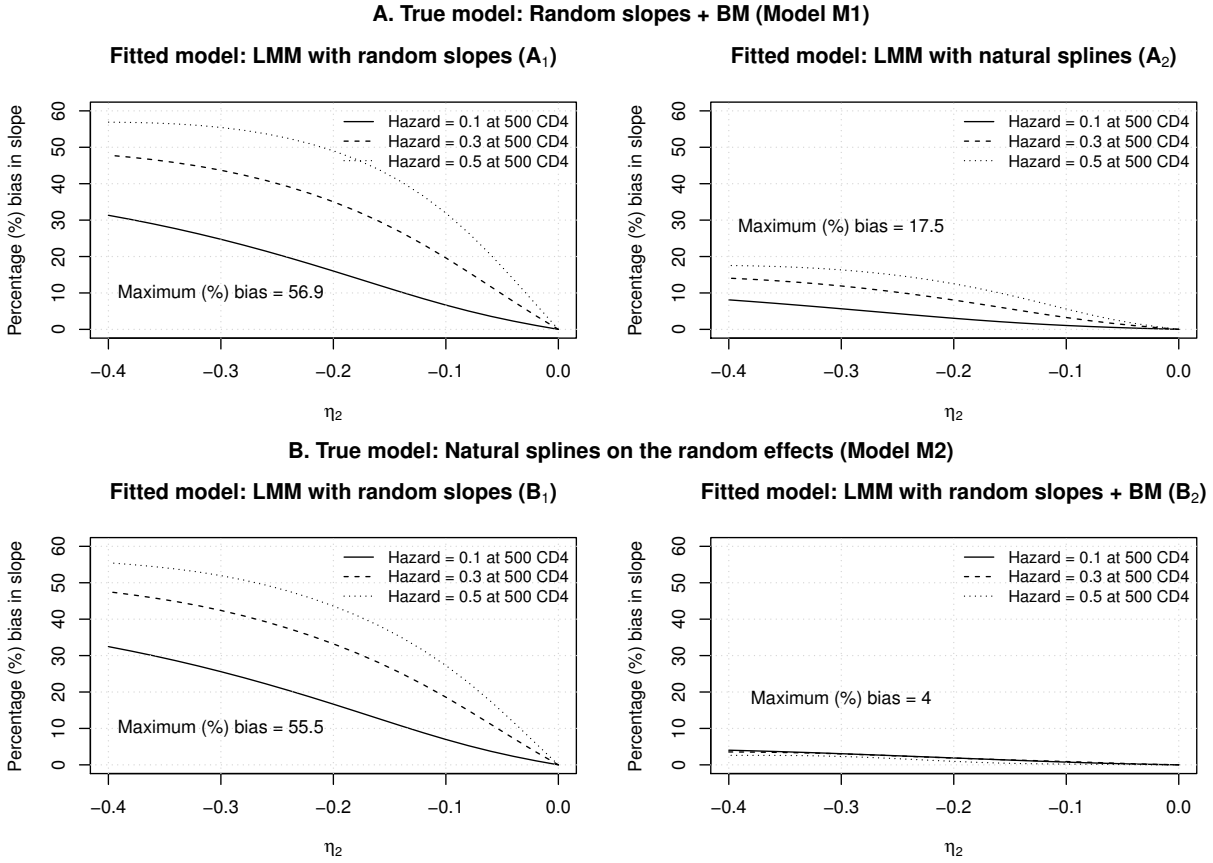


Figure 4.7: Asymptotic bias in the estimate of the ultimate slope of a marker (i.e. the rate of change after 1.42 years since baseline) under misspecified covariance structure and continuous-time dropout. **A**. True covariance structure based on an LMM with a fractional BM process on top of random intercepts and slopes (Model M1), while the fitted models are an LMM with random intercept and slope (Model M3), panel A_1, and an LMM with natural splines for the random effects (Model M2), panel A_2. **B**. True covariance structure described by M2 and fitted covariance structure by M3 (B_1) and M1 (B_2). The parameter η_2 measures the change in the log hazard rate of dropout associated with one unit decrease in the current marker value. Three scenarios regarding the hazard dropout rate at 500 CD4/μL, i.e. approximately the CD4 counts at seroconversion (baseline), were presented: 0.1, 0.30 and 0.50.

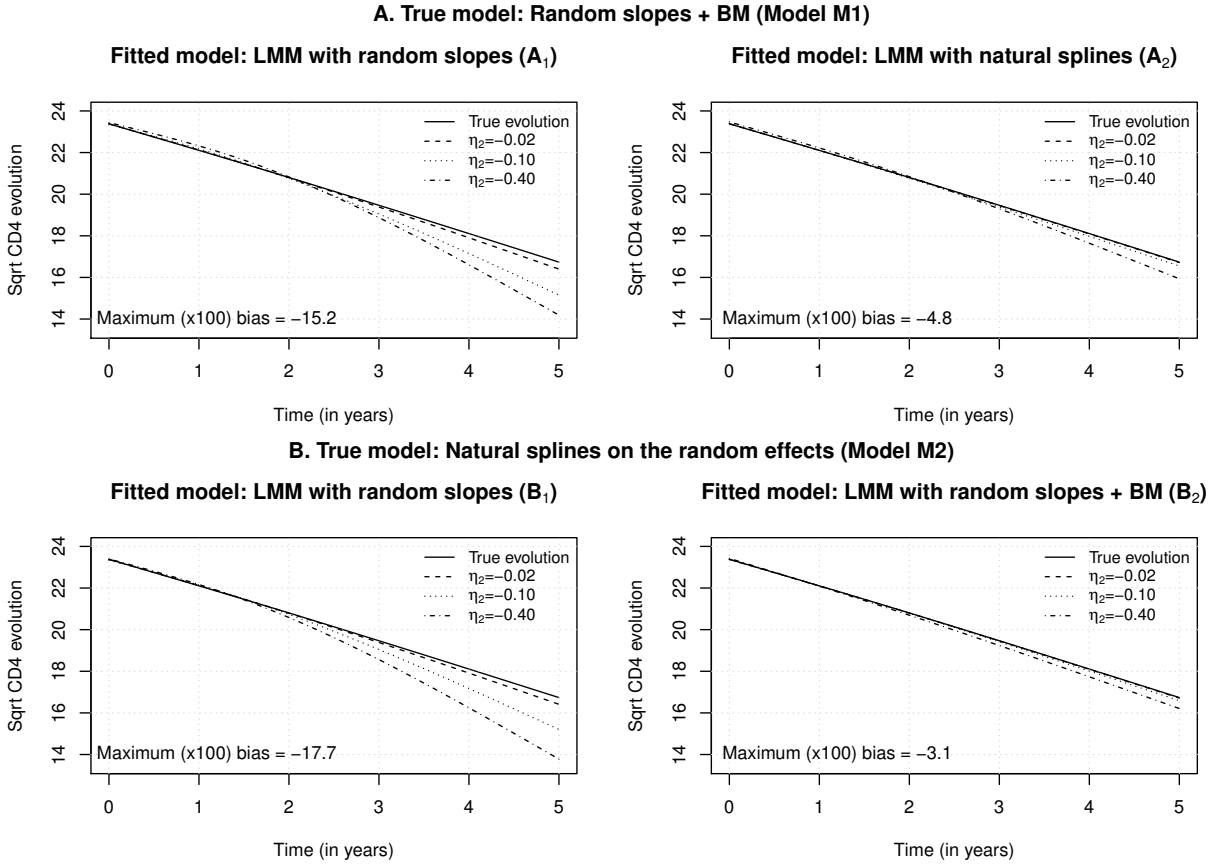


Figure 4.8: Asymptotic bias in a marker average evolution estimate under misspecified covariance structure and continuous-time dropout. **A**. True covariance structure based on an LMM with a fractional BM process on top of random intercepts and slopes (Model M1), while the fitted models are an LMM with random intercept and slope (Model M3), panel A_1, and an LMM with natural splines for the random effects (Model M2), panel A_2. **B**. True covariance structure described by M2 and fitted covariance structure by M3 (B_1) and M1 (B_2). The solid curve represents the "true" marker evolution, whereas the dashed, dotted and dot-dashed curves show the estimated marker evolution for η_2 equal to -0.02, -0.10, -0.40, respectively. The parameter η_2 measures the change in the log hazard rate of dropout associated with one unit decrease in the current marker value. The hazard of dropout at 500 CD4/μL, i.e. approximately the CD4 counts at seroconversion (baseline), was 0.50.

is equal to $(\beta_{G1}, \beta_{G2}, \beta_{G3})$, with the interest mainly lying on β_{G3}, i.e. the difference in the marker slope between the two groups. We also assume that the "true" covariance structure is the same across the two groups, with $\boldsymbol{\theta}_{L0,G}^{\top} = (\boldsymbol{\beta}_{0,G}^{\top}, \boldsymbol{\theta}_{v0,G}^{\top})$ denoting the "true" parameter values of the mean marker evolution and the variance components, respectively. As in Section 4.2, we assume that the mean evolution is correctly specified in the fitted model, whereas the covariance structure is misspecified. Let $\boldsymbol{\mu}_{0(j),g}$ and $\boldsymbol{\Omega}_{0(j),g}$ be the true mean and covariance matrix of $\boldsymbol{Y}_{(j)}|M = j, G = g; \boldsymbol{\theta}_{L0,G}, c_{1,g}, c_{2,g}$, and $a_{0j,g} = \Pr(M = j|G = g; \boldsymbol{\theta}_{L0,G}, c_{1,g}, c_{2,g})$ be the true dropout probabilities for group g, $g = 0, 1$. To derive the limit in probability of the misspecified model, we need to solve, with respect to $\boldsymbol{\theta}_{L,G}$, the following equation

$$\frac{\partial}{\partial \boldsymbol{\theta}_{L,G}} \sum_{g=0}^{1} \sum_{j=1}^{Q} p_g a_{0j,g} \Bigg[- \frac{\log |\sigma_G^2 \boldsymbol{V}_{(j),G}|}{2} - \frac{1}{2\sigma_G^2} \Bigg\{ \operatorname{tr}(\boldsymbol{V}_{(j),G}^{-1} \boldsymbol{\Omega}_{0(j),g}) + (\boldsymbol{\mu}_{0(j),g} - \boldsymbol{X}_{(j),g}\boldsymbol{\beta}_G)^{\top} \boldsymbol{V}_{(j),G}^{-1} (\boldsymbol{\mu}_{0(j),g} - \boldsymbol{X}_{(j),g}\boldsymbol{\beta}_G) \Bigg\} \Bigg] = \mathbf{0}, \tag{4.8}$$

where $\boldsymbol{\theta}_{L,G}^{\top} = (\boldsymbol{\beta}_G^{\top}, \boldsymbol{\theta}_{v,G}^{\top})$ denotes the parameters of the two-group misspecified model and $\sigma_G^2 \boldsymbol{V}_{(j),G}$ is the covariance matrix of the misspecified model for the jth dropout pattern. Recall that the covariance structure is assumed to be the same in the two groups.

4.4.1 A special case: Same Slope and Drop-out Mechanism in the Two Groups

We consider a special case in which the marker slope, as well as the dropout mechanisms, are the same in the two groups. Under these assumptions, it follows that $a_{0j,g} = a_{0j}$, $\boldsymbol{\mu}_{0(j),g} = \boldsymbol{\mu}_{0(j)}$, and $\boldsymbol{\Omega}_{0(j),g} = \boldsymbol{\Omega}_{0(j)}$, $g = 0, 1$. Thus, Equation (4.8) becomes equal to

$$\begin{aligned} &\frac{\partial}{\partial \boldsymbol{\theta}_{L,G}} \sum_{j=1}^{Q} \Bigg\{ - \frac{a_{0j} \log |\sigma_G^2 \boldsymbol{V}_{(j),G}|}{2} - \frac{a_{0j}}{2\sigma_G^2} \operatorname{tr}(\boldsymbol{V}_{(j),G}^{-1} \boldsymbol{\Omega}_{0(j)}) \Bigg\} \\ = &\frac{\partial}{\partial \boldsymbol{\theta}_{L,G}} \sum_{j=1}^{Q} \sum_{g=0}^{1} \Bigg\{ \frac{p_g a_{0j}}{2\sigma_G^2} (\boldsymbol{\mu}_{0(j)} - \boldsymbol{X}_{(j),g}\boldsymbol{\beta}_G)^{\top} \boldsymbol{V}_{(j),G}^{-1} (\boldsymbol{\mu}_{0(j)} - \boldsymbol{X}_{(j),g}\boldsymbol{\beta}_G) \Bigg\}, \end{aligned} \tag{4.9}$$

where the equality holds element by element. Let us also denote the solution to Equation (4.9) by ${\boldsymbol{\theta}^{\star}_{L,G}}^{\top} = ({\boldsymbol{\beta}^{\star}_{G}}^{\top}, {\boldsymbol{\theta}^{\star}_{v,G}}^{\top})$. Then the slope difference is consistently estimated if the third element of $\boldsymbol{\beta}^{\star}_{G}$ (i.e. $\beta^{\star}_{G3}$) is equal to zero; otherwise the estimate is biased.

Let ${\boldsymbol{\theta}^{\star}_{L}}^{\top} = (\boldsymbol{\beta}^{\star\top}, {\boldsymbol{\theta}^{\star}_{v}}^{\top})$ be the limit in probability of the estimates from the misspecified model assuming no time×group interaction. Note that $\boldsymbol{\beta}^{\star}$ is now of length two since it only includes the population-average constant and the time-slope. It easily follows that

$$\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{\theta}^{\star}_{L}} \sum_{j=1}^{Q} \left\{ -\frac{a_{0j} \log |\sigma^{\star 2} \boldsymbol{V}^{\star}_{(j)}|}{2} - \frac{a_{0j}}{2\sigma^{\star 2}} \mathrm{tr}({\boldsymbol{V}^{\star}_{(j)}}^{-1} \boldsymbol{\Omega}_{0(j)}) \right\} \\
&= \frac{\partial}{\partial \boldsymbol{\theta}^{\star}_{L}} \sum_{j=1}^{Q} \left\{ \frac{a_{0j}}{2\sigma^{\star 2}} (\boldsymbol{\mu}_{0(j)} - \boldsymbol{X}_{(j)} \boldsymbol{\beta}^{\star})^{\top} {\boldsymbol{V}^{\star}_{(j)}}^{-1} (\boldsymbol{\mu}_{0(j)} - \boldsymbol{X}_{(j)} \boldsymbol{\beta}^{\star}) \right\}, \qquad (4.10)
\end{aligned}$$

where $\boldsymbol{X}_{(j)}$ includes a column of ones and a column with the measurement times and $\sigma^{\star 2} \boldsymbol{V}^{\star}_{(j)}$ is the covariance matrix for the jth dropout pattern of the misspecified model assuming no time×group interaction at $\boldsymbol{\theta}^{\star}_{v}$; note that $\frac{\partial S(\boldsymbol{\theta}^{\star}_{L})}{\partial \boldsymbol{\theta}^{\star}_{L}}$ is interpreted to mean $\frac{\partial S(\boldsymbol{\theta}_{L})}{\partial \boldsymbol{\theta}_{L}}\big|_{\boldsymbol{\theta}_{L} = \boldsymbol{\theta}^{\star}_{L}}$.

We now show that the estimates from the misspecified model that assumes a time×group interaction converge to the same corresponding values as the model that does not include the interaction term, with the third element of $\boldsymbol{\beta}^{\star}_{G}$ being equal to zero, i.e. the slope difference is consistently estimated. To do so, we need to verify that $(\boldsymbol{\beta}^{\star\top}, 0)$ and $\boldsymbol{\theta}^{\star}_{v}$ solve Equation (4.9). Note that by first setting the third element of $\boldsymbol{\beta}_{G}$ (i.e. the slope difference parameter) equal to zero and evaluating the partial derivatives of Equation (4.9) for all the parameters except for the slope difference at ${\boldsymbol{\theta}^{\star}_{L}}^{\top} = (\boldsymbol{\beta}^{\star\top}, {\boldsymbol{\theta}^{\star}_{v}}^{\top})$, we end up with Equation (4.10), which holds by definition. This is evident upon observing that $\boldsymbol{X}_{(j),g}(\boldsymbol{\beta}^{\star\top}, 0)^{\top} = \boldsymbol{X}_{(j)} \boldsymbol{\beta}^{\star}$, for any $g = 0, 1$. Moreover, as it will be shown, evaluating Equation (4.9) at ${\boldsymbol{\theta}^{\star}_{L}}^{\top} = (\boldsymbol{\beta}^{\star\top}, {\boldsymbol{\theta}^{\star}_{v}}^{\top})$ and differentiating over β_{G3}, the slope difference parameter, the resulting solution over β_{G3} is zero, which proves that $(\boldsymbol{\beta}^{\star\top}, 0)$ and $\boldsymbol{\theta}^{\star}_{v}$ simultaneously solve Equation (4.9). Therefore, we have proved that if the true slope difference is zero and the dropout mechanism is the same in the two groups, the estimate of the slope difference from the misspecified model

is unbiased, irrespective of the covariance structure of the true data model and the parameters of the dropout mechanism.

To verify this result, we first set all the parameters except for the slope difference parameter equal to ${\boldsymbol{\theta}_L^\star}^\top = ({\boldsymbol{\beta}^\star}^\top, {\boldsymbol{\theta}_v^\star}^\top)$, i.e. the values to which the estimates from the model without time×group interaction converge, and then evaluate the derivatives over β_{G3}, resulting in

$$\frac{\partial}{\partial \beta_{G3}} \sum_{j=1}^{Q} \sum_{g=0}^{1} \frac{p_g a_{0j}}{2\sigma^{\star 2}} \left\{ \boldsymbol{\mu}_{0(j)} - \begin{pmatrix} \boldsymbol{X}_{(j)} & \boldsymbol{t}_{(j)} G \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}^\star \\ \beta_{G3} \end{pmatrix} \right\}^\top {\boldsymbol{V}_{(j)}^\star}^{-1} \left\{ \boldsymbol{\mu}_{0(j)} - \begin{pmatrix} \boldsymbol{X}_{(j)} & \boldsymbol{t}_{(j)} G \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}^\star \\ \beta_{G3} \end{pmatrix} \right\} = 0,$$

where $G = I(g = 1)$ is an indicator of Group 1, $\boldsymbol{X}_{(j)} = \begin{pmatrix} \mathbf{1} & \boldsymbol{t}_{(j)} \end{pmatrix}$, and $\sigma^{\star 2} \boldsymbol{V}_{(j)}^\star$ is the covariance matrix for the jth dropout pattern of the misspecifed model assuming no time×group interaction at $\boldsymbol{\theta}_v^\star$. It is thus implied that

$$\frac{\partial}{\partial \beta_{G3}} \sum_{j=1}^{Q} \sum_{g=0}^{1} \frac{p_g a_{0j}}{2\sigma^{\star 2}} (\boldsymbol{\mu}_{0(j)} - \boldsymbol{X}_{(j)}\boldsymbol{\beta}^\star - \beta_{G3} G \boldsymbol{t}_{(j)})^\top {\boldsymbol{V}_{(j)}^\star}^{-1} (\boldsymbol{\mu}_{0(j)} - \boldsymbol{X}_{(j)}\boldsymbol{\beta}^\star - \beta_{G3} G \boldsymbol{t}_{(j)}) = 0$$

$$\Rightarrow \frac{p_1}{2\sigma^{\star 2}} \frac{\partial}{\partial \beta_{G3}} \sum_{j=1}^{Q} a_{0j} \Big\{ (\boldsymbol{\mu}_{0(j)} - \boldsymbol{X}_{(j)}\boldsymbol{\beta}^\star)^\top {\boldsymbol{V}_{(j)}^\star}^{-1} (\boldsymbol{\mu}_{0(j)} - \boldsymbol{X}_{(j)}\boldsymbol{\beta}^\star) + \beta_{G3}^2 \boldsymbol{t}_{(j)}^\top {\boldsymbol{V}_{(j)}^\star}^{-1} \boldsymbol{t}_{(j)}$$

$$- 2\beta_{G3} \boldsymbol{t}_{(j)}^\top {\boldsymbol{V}_{(j)}^\star}^{-1} (\boldsymbol{\mu}_{0(j)} - \boldsymbol{X}_{(j)}\boldsymbol{\beta}^\star) \Big\} = 0$$

$$\Rightarrow \sum_{j=1}^{Q} a_{0j} \left\{ 2\beta_{G3} \boldsymbol{t}_{(j)}^\top {\boldsymbol{V}_{(j)}^\star}^{-1} \boldsymbol{t}_{(j)} - 2\boldsymbol{t}_{(j)}^\top {\boldsymbol{V}_{(j)}^\star}^{-1} (\boldsymbol{\mu}_{0(j)} - \boldsymbol{X}_{(j)}\boldsymbol{\beta}^\star) \right\} = 0$$

$$\Rightarrow \beta_{G3}^\star = \left(\sum_{j=1}^{Q} a_{0j} \boldsymbol{t}_{(j)}^\top {\boldsymbol{V}_{(j)}^\star}^{-1} \boldsymbol{t}_{(j)} \right)^{-1} \sum_{j=1}^{Q} a_{0j} \boldsymbol{t}_{(j)}^\top {\boldsymbol{V}_{(j)}^\star}^{-1} (\boldsymbol{\mu}_{0(j)} - \boldsymbol{X}_{(j)}\boldsymbol{\beta}^\star), \tag{4.11}$$

where $\beta_{G3}^\star$ is the value to which the estimate of the slope difference converges. Taking into account that $(\boldsymbol{\beta}^\star, \boldsymbol{\theta}_v^\star)$ solve Equation (4.10), it follows that

$$\frac{\partial}{\partial \boldsymbol{\beta}^\star} \sum_{j=1}^{Q} \left\{ \frac{a_{0j}}{2\sigma^{\star 2}} (\boldsymbol{\mu}_{0(j)}^\top {\boldsymbol{V}_{(j)}^\star}^{-1} \boldsymbol{\mu}_{0(j)} + {\boldsymbol{\beta}^\star}^\top \boldsymbol{X}_{(j)}^\top {\boldsymbol{V}_{(j)}^\star}^{-1} \boldsymbol{X}_{(j)} \boldsymbol{\beta}^\star - 2{\boldsymbol{\beta}^\star}^\top \boldsymbol{X}_{(j)}^\top {\boldsymbol{V}_{(j)}^\star}^{-1} \boldsymbol{\mu}_{0(j)}) \right\} = \mathbf{0},$$

which further implies that

$$\begin{aligned}
\sum_{j=1}^{Q} a_{0j} \boldsymbol{X}_{(j)}^{\top} {\boldsymbol{V}_{(j)}^{\star}}^{-1} (\boldsymbol{\mu}_{0(j)} - \boldsymbol{X}_{(j)} \boldsymbol{\beta}^{\star}) &= \mathbf{0} \\
\Rightarrow \sum_{j=1}^{Q} a_{0j} \begin{pmatrix} \mathbf{1}^{\top} \\ \boldsymbol{t}_{(j)}^{\top} \end{pmatrix} {\boldsymbol{V}_{(j)}^{\star}}^{-1} (\boldsymbol{\mu}_{0(j)} - \boldsymbol{X}_{(j)} \boldsymbol{\beta}^{\star}) &= \mathbf{0} \\
\Rightarrow \begin{pmatrix} \sum_{j=1}^{Q} a_{0j} \mathbf{1}^{\top} {\boldsymbol{V}_{(j)}^{\star}}^{-1} (\boldsymbol{\mu}_{0(j)} - \boldsymbol{X}_{(j)} \boldsymbol{\beta}^{\star}) \\ \sum_{j=1}^{Q} a_{0j} \boldsymbol{t}_{(j)}^{\top} {\boldsymbol{V}_{(j)}^{\star}}^{-1} (\boldsymbol{\mu}_{0(j)} - \boldsymbol{X}_{(j)} \boldsymbol{\beta}^{\star}) \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix}.
\end{aligned}$$

Therefore, we have shown that $\sum_{j=1}^{Q} a_{0j} \boldsymbol{t}_{(j)}^{\top} {\boldsymbol{V}_{(j)}^{\star}}^{-1} (\boldsymbol{\mu}_{0(j)} - \boldsymbol{X}_{(j)} \boldsymbol{\beta}^{\star}) = 0$, which, based on Equation (4.11), proves that $\beta_{G3}^{\star}$ is equal to zero.

4.4.2 Numerical Evaluation of the Bias in the Slope Difference Estimate

We numerically quantified the bias in the slope difference estimates, considering the models described in Subsection 4.2.1. The "true" slope difference was assumed to be 0; i.e. $\boldsymbol{\beta}_{0,G} = (23.24, -1.12, 0)^{\top}$, whereas the dropout mechanisms applied to both groups were allowed to be different, assuming though common baseline dropout rate, i.e. $P(M = j | M \geq j, \boldsymbol{Y}_{(j),g}, G = g; c_1, c_{2,g}) = \frac{\exp\{c_1 + c_{2,g}(Y_{j,g} - Y^{\star})\}}{1 + \exp\{c_1 + c_{2,g}(Y_{j,g} - Y^{\star})\}}$. For group 0 we assumed that $c_{2,0} = -0.24$, whereas for group 1 we varied $c_{2,1}$ from 0 to -0.40. The results are presented in Figure 4.9. It is evident that the bias is exactly equal to zero when $c_{2,1} = c_{2,0}$, i.e. the two groups have the same dropout mechanism, confirming our analytical findings. The bias increased as the difference between $c_{2,1}$ and $c_{2,0}$ increased, i.e. it became higher when the two dropout mechanisms were substantially different. Furthermore, similarly to the results in Section 4.2, the highest biases in the estimated slope difference were obtained by model M3 (absolute bias up to 0.16). The models M1 and M2 performed similarly well, though the bias in the estimated slope difference was slightly lower in the former (up to 0.03) compared to the latter (up to 0.06).

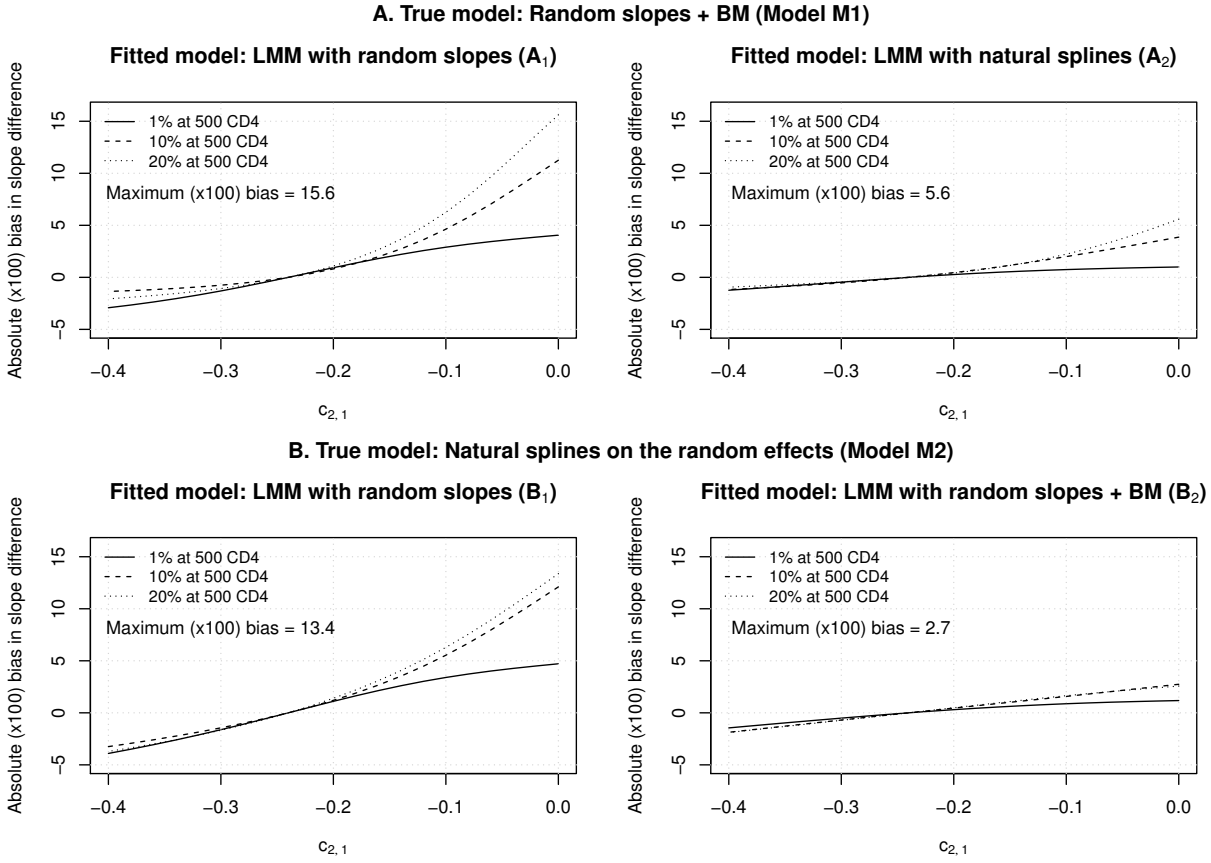


Figure 4.9: Asymptotic bias in the marker slope difference estimate between two groups under misspecified covariance structure. **A**. True covariance structure based on an LMM with a fractional BM process on top of random intercepts and slopes (Model M1), while the fitted models are an LMM with random intercept and slope (Model M3), panel A_1, and an LMM with natural splines for the random effects (Model M2), panel A_2. **B**. True covariance structure described by M2 and fitted covariance structure by M3 (B_1) and M1 (B_2). The parameter $c_{2,1}$ measures the change in the log-odds of dropout associated with one unit decrease in the current marker value for group 1, whereas the corresponding value for group 0 is $c_{2,0} = -0.24$. Three scenarios regarding the common dropout rate at 500 CD4/μL, i.e. approximately the CD4 counts at seroconversion (baseline), were presented: 1%, 10% and 20%.

4.5 Bayesian model comparison

To compare two non-nested models, i.e. LMMs with different covariance structures, we adopt a fully Bayesian model comparison approach based on the probability of a given model being true given the data, i.e. the posterior model probability (PMP). To refer to subjects, we reintroduce subscript i assuming also that the visit times may differ between subjects. Recall that, given $M_i = m_i$, the observed marker measurements on subject i are $\boldsymbol{Y}_{i,(m_i)}^{\top} = (Y_{i1}, Y_{i2}, \ldots, Y_{im_i})$, collected at times $t_{i1}, \ldots, t_{im_i}$. We aim to compare the fit of the models M1 and M2. Letting $\boldsymbol{\theta}_{L1}$ be the whole parameter vector of M1 and N the number of subjects, the marginal likelihood of M1 is $f(\boldsymbol{Y}_{1,(m_1)}, \ldots, \boldsymbol{Y}_{N,(m_N)}|\text{M1}) = \int \prod_{i=1}^{N} f(\boldsymbol{Y}_{i,(m_i)}|\text{M1}; \boldsymbol{\theta}_{L1}) f(\boldsymbol{\theta}_{L1}|\text{M1}) d\boldsymbol{\theta}_{L1}$, i.e. the marginal posterior probability of the data given the model or, equivalently, the expectation of the likelihood over the prior. Assuming equal prior probabilities on the two models, the posterior model probability (PMP) of M1 is equal to

$$P(\text{M1}|\boldsymbol{Y}_{1,(m_1)}, \ldots, \boldsymbol{Y}_{N,(m_N)}) = \frac{f(\boldsymbol{Y}_{1,(m_1)}, \ldots, \boldsymbol{Y}_{N,(m_N)}|\text{M1})}{f(\boldsymbol{Y}_{1,(m_1)}, \ldots, \boldsymbol{Y}_{N,(m_N)}|\text{M1}) + f(\boldsymbol{Y}_{1,(m_1)}, \ldots, \boldsymbol{Y}_{N,(m_N)}|\text{M2})}.$$

In the case of comparing K competitive models, the PMP of M_j, $j = 1, 2, \ldots, K$, is given by $f(\boldsymbol{Y}_{1,(m_1)}, \ldots, \boldsymbol{Y}_{N,(m_N)}|M_j) / \sum_{k=1}^{K} f(\boldsymbol{Y}_{1,(m_1)}, \ldots, \boldsymbol{Y}_{N,(m_N)}|M_k)$. Then the model with the highest posterior probability is preferred.

For comparing M1 and M2, we assume the following prior distributions; $\omega = 1/\sigma^2 \sim$ Gamma(a_0, λ_0) and $\boldsymbol{\beta}|\omega \sim N(\boldsymbol{\mu}_0, \omega^{-1}\boldsymbol{C}_0)$. In M1 we assume a $N(\mu_\kappa, \sigma_\kappa^2)$ distribution truncated on $(0, \infty)$ and a Beta(α_1, α_2) distribution for the parameters κ and H of the fractional BM process, respectively. As it will be shown, $\boldsymbol{\beta}$ and ω can be analytically integrated out (integration over $\boldsymbol{\beta}$ is equivalent to the method used to obtain the restricted maximum likelihood estimates for LMMs), with the marginal likelihood being

$$\int c(\boldsymbol{\theta}_v)\Gamma\left(\frac{n}{2} + a_0\right)\left(\lambda_0 + \frac{\sum_{i=1}^{N} \boldsymbol{Y}_{i,(m_i)}^{\top} \boldsymbol{V}_{i,(m_i)}^{-1} \boldsymbol{Y}_{i,(m_i)}}{2} + \frac{\boldsymbol{\mu}_0^{\top} \boldsymbol{C}_0^{-1} \boldsymbol{\mu}_0}{2} - \frac{\boldsymbol{\mu}_1^{\top} \boldsymbol{C}_1^{-1} \boldsymbol{\mu}_1}{2}\right)^{-(\frac{n}{2}+a_0)} d\boldsymbol{\theta}_v, \tag{4.12}$$

where $n = \sum_{i=1}^{N} m_i$, and $\boldsymbol{V}_{i,(m_i)}$ is equal to $(\boldsymbol{I}_{(m_i)} + \boldsymbol{Z}_{1i,(m_i)} \boldsymbol{D}_1 \boldsymbol{Z}_{1i,(m_i)}^{\top} + \boldsymbol{\Sigma}_{i,(m_i)})$ and $(\boldsymbol{I}_{(m_i)} + \boldsymbol{Z}_{2i,(m_i)} \boldsymbol{D}_2 \boldsymbol{Z}_{2i,(m_i)}^{\top})$ for models M1 and M2, respectively. Recall that $\boldsymbol{\Sigma}_{i,(m_i)}$ is

associated with the covariance matrix of the fractional BM process. Also, $\boldsymbol{C}_1$ and $\boldsymbol{\mu}_1$ are equal to $(\boldsymbol{C}_0^{-1}+\sum_{i=1}^{N}\boldsymbol{X}_{i,(m_i)}^{\top}\boldsymbol{V}_{i,(m_i)}^{-1}\boldsymbol{X}_{i,(m_i)})^{-1}$ and $\boldsymbol{C}_1(\boldsymbol{C}_0^{-1}\boldsymbol{\mu}_0+\sum_{i=1}^{N}\boldsymbol{X}_{i,(m_i)}^{\top}\boldsymbol{V}_{i,(m_i)}^{-1}\boldsymbol{Y}_{i,(m_i)})$, respectively, while $c(\boldsymbol{\theta}_v)$ is equal to $(2\pi)^{-n/2}\prod_{i=1}^{N}|\boldsymbol{V}_{i,(m_i)}|^{-1/2}\lambda_0^{a_0}|\boldsymbol{C}_0|^{-1/2}|\boldsymbol{C}_1|^{1/2}f(\boldsymbol{\theta}_v)/\Gamma(a_0)$, where $f(\boldsymbol{\theta}_v)$ is the prior distribution on $\boldsymbol{\theta}_v$.

We first derive the marginal likelihood of the M1 model; for M2, the process is very similar. Recall that the form of the model is $\boldsymbol{Y}_{i,(m_i)} = \boldsymbol{X}_{i,(m_i)}\boldsymbol{\beta}_1 + \boldsymbol{Z}_{1i,(m_i)}\boldsymbol{b}_{1i} + \boldsymbol{W}_{i,(m_i)} + \boldsymbol{\epsilon}_{1i,(m_i)}$, where $\boldsymbol{b}_{1i} \sim N(\boldsymbol{0}, \omega_1^{-1}\boldsymbol{D}_1)$ is the vector of the random effects and $\boldsymbol{\epsilon}_{1i,(m_i)} \sim N(\boldsymbol{0}, \omega_1^{-1}\boldsymbol{I}_{(m_i)})$ is the vector of the within-subjects errors, assumed to be independent of one another. Also, ω_1 denotes the within-subjects precision, i.e. the inverse of the within-subjects variance in the M1 model. We denote the fractional BM process defined at the measurement times of the ith subject $t_{i1}, \ldots, t_{im_i}$ by $\boldsymbol{W}_{i,(m_i)}$, which implies that $\boldsymbol{W}_{i,(m_i)} \sim N(\boldsymbol{0}, \omega_1^{-1}\boldsymbol{\Sigma}_{i,(m_i)})$, where $\boldsymbol{\Sigma}_{i,(m_i)}$ is parametrized in terms of two parameters, κ and H. With these definitions and $\boldsymbol{V}_{i,(m_i)} = \boldsymbol{I}_{(m_i)}+\boldsymbol{Z}_{1i,(m_i)}\boldsymbol{D}_1\boldsymbol{Z}_{1i,(m_i)}^{\top}+\boldsymbol{\Sigma}_{i,(m_i)}$, the observed likelihood of the ith subject is equal to

$$
\begin{aligned}
&f(\boldsymbol{Y}_{i,(m_i)};\boldsymbol{\theta}_{L1}) = (2\pi)^{-m_i/2}|\omega_1^{-1}\boldsymbol{V}_{i,(m_i)}|^{-1/2} \\
&\times \exp\left\{-\frac{\omega_1}{2}(\boldsymbol{Y}_{i,(m_i)} - \boldsymbol{X}_{i,(m_i)}\boldsymbol{\beta}_1)^{\top}\boldsymbol{V}_{i,(m_i)}^{-1}(\boldsymbol{Y}_{i,(m_i)} - \boldsymbol{X}_{i,(m_i)}\boldsymbol{\beta}_1)\right\} \\
&= (2\pi)^{-m_i/2}\omega_1^{m_i/2}|\boldsymbol{V}_{i,(m_i)}|^{-1/2} \times \exp\left\{ -\frac{\omega_1}{2}\Big(\boldsymbol{Y}_{i,(m_i)}^{\top}\boldsymbol{V}_{i,(m_i)}^{-1}\boldsymbol{Y}_{i,(m_i)}\right. \\
&\left.-2\boldsymbol{\beta}_1^{\top}\boldsymbol{X}_{i,(m_i)}^{\top}\boldsymbol{V}_{i,(m_i)}^{-1}\boldsymbol{Y}_{i,(m_i)} + \boldsymbol{\beta}_1^{\top}\boldsymbol{X}_{i,(m_i)}^{\top}\boldsymbol{V}_{i,(m_i)}^{-1}\boldsymbol{X}_{i,(m_i)}\boldsymbol{\beta}_1\Big)\right\}.
\end{aligned}
$$

Therefore, letting $n = \sum_{i=1}^{N} m_i$, where N denotes the number of subjects, the full likelihood is equal to

$$
\begin{aligned}
&f(\boldsymbol{Y}_{1,(m_1)}, \ldots, \boldsymbol{Y}_{N,(m_N)};\boldsymbol{\theta}_{L1}) = (2\pi)^{-n/2}\omega_1^{n/2}\prod_{i=1}^{N}|\boldsymbol{V}_{i,(m_i)}|^{-1/2} \\
&\times \exp\left\{ -\frac{\omega_1}{2}\sum_{i=1}^{N}\Big(\boldsymbol{Y}_{i,(m_i)}^{\top}\boldsymbol{V}_{i,(m_i)}^{-1}\boldsymbol{Y}_{i,(m_i)}\right. \\
&\left.-2\boldsymbol{\beta}_1^{\top}\boldsymbol{X}_{i,(m_i)}^{\top}\boldsymbol{V}_{i,(m_i)}^{-1}\boldsymbol{Y}_{i,(m_i)} + \boldsymbol{\beta}_1^{\top}\boldsymbol{X}_{i,(m_i)}^{\top}\boldsymbol{V}_{i,(m_i)}^{-1}\boldsymbol{X}_{i,(m_i)}\boldsymbol{\beta}_1\Big)\right\}.
\end{aligned}
$$

Assuming a $N(\boldsymbol{\mu}_0, \omega_1^{-1}\boldsymbol{C}_0)$ distribution for the fixed effects

$$f(\boldsymbol{\beta}_1|\omega_1) = (2\pi)^{-p/2}\omega_1^{p/2}|\boldsymbol{C}_0|^{-1/2}\exp\left\{-\frac{\omega_1}{2}(\boldsymbol{\beta}_1^\top \boldsymbol{C}_0{}^{-1}\boldsymbol{\beta}_1 - 2\boldsymbol{\beta}_1^\top \boldsymbol{C}_0{}^{-1}\boldsymbol{\mu}_0 + \boldsymbol{\mu}_0^\top \boldsymbol{C}_0{}^{-1}\boldsymbol{\mu}_0)\right\}$$

as well as a Gamma distribution for ω_1, $f(\omega_1) = \frac{\lambda_0^{\alpha_0}}{\Gamma(\alpha_0)}\omega_1^{\alpha_0-1}\exp(-\lambda_0\omega_1)$, and integrating the likelihood over the prior distribution of the fixed effects results in

$$\begin{aligned}
&\int f(\boldsymbol{Y}_{1,(m_1)},\ldots,\boldsymbol{Y}_{N,(m_N)};\boldsymbol{\theta}_{L1})f(\boldsymbol{\beta}_1|\omega_1)f(\omega_1)f(\boldsymbol{\theta}_{v1})d\boldsymbol{\beta}_1 \\
= \;& (2\pi)^{-\frac{n+p}{2}}\prod_{i=1}^{N}|\boldsymbol{V}_{i,(m_i)}|^{-1/2}\frac{\lambda_0^{\alpha_0}}{\Gamma(\alpha_0)}f(\boldsymbol{\theta}_{v1})|\boldsymbol{C}_0|^{-1/2}\omega_1^{\frac{n+p}{2}+\alpha_0-1} \\
\times \;& \exp\left\{-\Big(\lambda_0 + \frac{1}{2}\sum_{i=1}^{N}\boldsymbol{Y}_{i,(m_i)}^\top \boldsymbol{V}_{i,(m_i)}^{-1}\boldsymbol{Y}_{i,(m_i)}\right. \\
+ \;& \left.\frac{1}{2}\boldsymbol{\mu}_0^\top \boldsymbol{C}_0{}^{-1}\boldsymbol{\mu}_0\Big)\omega_1\right\}\int \exp\left[-\frac{\omega_1}{2}\left\{\boldsymbol{\beta}_1^\top\left(\boldsymbol{C}_0{}^{-1} + \sum_{i=1}^{N}\boldsymbol{X}_{i,(m_i)}^\top \boldsymbol{V}_{i,(m_i)}^{-1}\boldsymbol{X}_{i,(m_i)}\right)\boldsymbol{\beta}_1\right.\right. \\
- \;& \left.\left.2\boldsymbol{\beta}_1^\top\left(\boldsymbol{C}_0{}^{-1}\boldsymbol{\mu}_0 + \sum_{i=1}^{N}\boldsymbol{X}_{i,(m_i)}^\top \boldsymbol{V}_{i,(m_i)}^{-1}\boldsymbol{Y}_{i,(m_i)}\right)\right\}\right]d\boldsymbol{\beta}_1. \qquad (4.13)
\end{aligned}$$

For notation simplicity, let

$$\begin{aligned}
\boldsymbol{C}_1{}^{-1} &= \left(\boldsymbol{C}_0{}^{-1} + \sum_{i=1}^{N}\boldsymbol{X}_{i,(m_i)}^\top \boldsymbol{V}_{i,(m_i)}^{-1}\boldsymbol{X}_{i,(m_i)}\right) \\
\boldsymbol{\mu}_1 &= \boldsymbol{C}_1\left(\boldsymbol{C}_0{}^{-1}\boldsymbol{\mu}_0 + \sum_{i=1}^{N}\boldsymbol{X}_{i,(m_i)}^\top \boldsymbol{V}_{i,(m_i)}^{-1}\boldsymbol{Y}_{i,(m_i)}\right).
\end{aligned}$$

After completing the probability density function of the $N(\boldsymbol{\mu}_1, \omega_1^{-1}\boldsymbol{C}_1)$, it can be shown that Equation (4.13) is equal to

$$\begin{aligned}
&(2\pi)^{-n/2}\prod_{i=1}^{N}|\boldsymbol{V}_{i,(m_i)}|^{-1/2}\frac{\lambda_0^{\alpha_0}}{\Gamma(\alpha_0)}f(\boldsymbol{\theta}_{v1})|\boldsymbol{C}_0|^{-1/2}|\boldsymbol{C}_1|^{1/2}\omega_1^{n/2+\alpha_0-1}\times\exp\left\{-\Big(\lambda_0\right. \\
+ \;& \left.\frac{1}{2}\sum_{i=1}^{N}\boldsymbol{Y}_{i,(m_i)}^\top \boldsymbol{V}_{i,(m_i)}^{-1}\boldsymbol{Y}_{i,(m_i)} + \frac{1}{2}\boldsymbol{\mu}_0^\top \boldsymbol{C}_0{}^{-1}\boldsymbol{\mu}_0 - \frac{1}{2}\boldsymbol{\mu}_1^\top \boldsymbol{C}_1{}^{-1}\boldsymbol{\mu}_1\Big)\omega_1\right\}. \qquad (4.14)
\end{aligned}$$

Integrating also ω_1 out of Equation (4.14), we have that

$$\begin{aligned}
& \int f(\boldsymbol{Y}_{1,(m_1)},\ldots,\boldsymbol{Y}_{N,(m_N)};\boldsymbol{\theta}_{L1})f(\boldsymbol{\beta}_1|\omega_1)f(\omega_1)f(\boldsymbol{\theta}_{v1})d\boldsymbol{\beta}_1 d\omega_1 \\
= \;& (2\pi)^{-n/2}\prod_{i=1}^{N}|\boldsymbol{V}_{i,(m_i)}|^{-1/2}\frac{\lambda_0^{\alpha_0}}{\Gamma(\alpha_0)}f(\boldsymbol{\theta}_{v1})|\boldsymbol{C}_0|^{-1/2}|\boldsymbol{C}_1|^{1/2}\int \omega_1^{n/2+\alpha_0-1}\exp\Bigg\{-\Big(\lambda_0 \\
+ \;& \frac{1}{2}\sum_{i=1}^{N}\boldsymbol{Y}_{i,(m_i)}^{\top}\boldsymbol{V}_{i,(m_i)}^{-1}\boldsymbol{Y}_{i,(m_i)}+\frac{1}{2}\boldsymbol{\mu}_0^{\top}\boldsymbol{C}_0{}^{-1}\boldsymbol{\mu}_0-\frac{1}{2}\boldsymbol{\mu}_1^{\top}\boldsymbol{C}_1{}^{-1}\boldsymbol{\mu}_1\Big)\omega_1\Bigg\}d\omega \\
= \;& (2\pi)^{-n/2}\prod_{i=1}^{N}|\boldsymbol{V}_{i,(m_i)}|^{-1/2}\frac{\lambda_0^{\alpha_0}}{\Gamma(\alpha_0)}f(\boldsymbol{\theta}_{v1})|\boldsymbol{C}_0|^{-1/2}|\boldsymbol{C}_1|^{1/2} \\
\times \;& \Gamma(n/2+\alpha_0)\Bigg/\left(\lambda_0+\frac{1}{2}\sum_{i=1}^{N}\boldsymbol{Y}_{i,(m_i)}^{\top}\boldsymbol{V}_{i,(m_i)}^{-1}\boldsymbol{Y}_{i,(m_i)}+\frac{1}{2}\boldsymbol{\mu}_0^{\top}\boldsymbol{C}_0{}^{-1}\boldsymbol{\mu}_0-\frac{1}{2}\boldsymbol{\mu}_1^{\top}\boldsymbol{C}_1{}^{-1}\boldsymbol{\mu}_1\right)^{n/2+\alpha_0} \qquad (4.15)
\end{aligned}$$

When the LMM with splines for the random effects is used, the only difference is that $\boldsymbol{\Sigma}_{i,(m_i)}$ is missing, i.e. $\boldsymbol{V}_{i,(m_i)} = \boldsymbol{I}_{(m_i)} + \boldsymbol{Z}_{2i,(m_i)}\boldsymbol{D}_2\boldsymbol{Z}_{2i,(m_i)}^{\top}$. To also integrate out the variance components (i.e. $\boldsymbol{\theta}_v$), we used adaptive quadrature for the BM model and quasi-Monte carlo integration for the LMM with natural splines for the random effects. In summary, both the fixed effects and the within-subjects precision are explicitly integrated out, which makes our method feasible irrespective of the dimension of the fixed effects. Instead, the variance components, the dimension of which is usually expected to be up to 5 or 6, are integrated out through numerical methods.

Among all prior specifications, the prior distribution on $\boldsymbol{D}$ needs the greatest care as it may involve a large number of parameters, which are also constrained such that the resulting matrix $\boldsymbol{D}$ to be positive definite. To deal with this issue, we applied the Inverse-Wishart $IW(\boldsymbol{A},\nu)$ prior distribution for $\boldsymbol{D}$, which has the advantage of being very popular in real data applications while easy to use, as it only requires the specification of $\boldsymbol{A}$ and the associated degrees of freedom, ν. It has been suggested that a good default specification is to take $\boldsymbol{A}$ to be equal to a best guess at $\boldsymbol{D}$ (based on literature and/or historical data from past studies) times the degrees of freedom, ν (Natarajan and Kass, 2000). For the degrees of freedom, a small value greater than $q-1$ is generally suggested, since the variance is non-decreasing in ν.

As an alternative prior specification, we also considered the separation strategy

approach proposed by Barnard et al. (2000), which factorized $\boldsymbol{D}$ as diag($\boldsymbol{S}$)$\boldsymbol{R}$diag($\boldsymbol{S}$), where $\boldsymbol{S}$ is a $q \times 1$ vector of standard deviations, diag($\boldsymbol{S}$) is a diagonal matrix with diagonal elements $\boldsymbol{S}$, and $\boldsymbol{R}$ is the $q \times q$ correlation matrix. Based on Barnard et al. (2000), we assumed a prior distribution for $\boldsymbol{R}$ that has marginally uniform pairwise correlations and independent truncated Normal distributions for the variances in $\boldsymbol{D}$.

Equation (4.12) can be approximated using numerical methods. For example, when the number of the random effects is relatively small (e.g. in M1 which involves 5 parameters to be integrated out), the adaptive quadrature procedure can be efficiently applied to approximate the integrals after centering and scaling the integral based on the posterior mode and the inverse curvature at the mode (Pinheiro and Bates, 1995). If the number of the random effects is large, the number of points at which the marginal likelihood is to be computed grows exponentially, which can make the use of the adaptive quadrature method computationally inefficient. In such cases, we can follow quasi Monte Carlo integration (Morokoff and Caflisch, 1995) coupled with importance sampling using a prespecified number of draws, taking the importance density to be multivariate normal with mean equal to the posterior mode and covariance matrix equal to the inverse curvature at the posterior mode (Pinheiro and Bates, 1995).

However, the previous methods require an unrestricted parameter space. To do so, we can reparameterize $\boldsymbol{D}$ in terms of the unique elements of its matrix logarithm (Leonard and Hsu, 1992) and calculate the resulting prior distribution on the matrix logarithm scale. However, this requires the Jacobian of the transformation, which is described below.

Jacobian of a Transformation to the Matrix Logarithm of $\boldsymbol{D}$

Firstly, we assume that $\boldsymbol{D} \sim IW(\boldsymbol{A}, \nu)$. To obtain the probability density of the matrix logarithm (Leonard and Hsu, 1992) of $\boldsymbol{D}$, we need to derive the Jacobian $|\frac{\partial \operatorname{vech}(e^{\boldsymbol{Y}})}{\partial \operatorname{vech}(\boldsymbol{Y})^\top}|$, where vech stands for the "vector-half" operation, $\boldsymbol{Y}$ is the matrix logarithm of $\boldsymbol{D}$ and $e^{\boldsymbol{Y}}$ the matrix exponential of $\boldsymbol{Y}$. To simplify the calculations, it is often much easier to work with the "Vec" operator, i.e. $|\frac{\partial \operatorname{vec}(e^{\boldsymbol{Y}})}{\partial \operatorname{vec}(\boldsymbol{Y})^\top}|$. It is well known that one efficient way to

compute the matrix exponential $e^{\boldsymbol{Y}}$ is through the eigenvalue-eigenvector decomposition. Writing $\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{\Delta}\boldsymbol{U}^\top$, where $\boldsymbol{U} = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_q]$ is the orthogonal matrix including the normalized eigenvectors (i.e. $\boldsymbol{u}_j^\top \boldsymbol{u}_j = 1$) and $\boldsymbol{\Delta}$ is the diagonal matrix with the respective eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_q$ on the main diagonal, then $e^{\boldsymbol{Y}} = \boldsymbol{U} e^{\boldsymbol{\Delta}} \boldsymbol{U}^\top$, where $e^{\boldsymbol{\Delta}}$ is still a diagonal matrix, including $e^{\lambda_1}, e^{\lambda_2}, \ldots, e^{\lambda_q}$ on the main diagonal. By the product rule of differentiation and the definition of the commutation matrix Harville (1997), i.e. $\operatorname{vec}(\boldsymbol{D}^\top) = \boldsymbol{K}_{qq} \operatorname{vec}(\boldsymbol{D})$, it is evident that

$$\frac{\partial \operatorname{vec}(e^{\boldsymbol{Y}})}{\partial \operatorname{vec}(\boldsymbol{Y})^\top} = (\boldsymbol{U}\boldsymbol{\Delta} \otimes \boldsymbol{I}) \frac{\partial \operatorname{vec}(\boldsymbol{U})}{\partial \operatorname{vec}(\boldsymbol{Y})^\top} + (\boldsymbol{U} \otimes \boldsymbol{U}) \frac{\partial \operatorname{vec}(e^{\boldsymbol{\Delta}})}{\partial \operatorname{vec}(\boldsymbol{Y})^\top} + (\boldsymbol{I} \otimes \boldsymbol{U}\boldsymbol{\Delta}) \boldsymbol{K}_{qq} \frac{\partial \operatorname{vec}(\boldsymbol{U})}{\partial \operatorname{vec}(\boldsymbol{Y})^\top}. \tag{4.16}$$

Thus, we need to derive $\frac{\partial \operatorname{vec}(\boldsymbol{U})}{\partial \operatorname{vec}(\boldsymbol{Y})^\top}$ and $\frac{\partial \operatorname{vec}(e^{\boldsymbol{\Delta}})}{\partial \operatorname{vec}(\boldsymbol{Y})^\top}$. As shown by Magnus (1985), $\frac{\partial \boldsymbol{u}_j}{\partial \operatorname{vec}(\boldsymbol{Y})^\top} = \boldsymbol{u}_j^\top \otimes (\lambda_j \boldsymbol{I} - \boldsymbol{Y})^+$, where $\boldsymbol{Y}^+$ denotes the Moore-Penrose inverse of $\boldsymbol{Y}$. Since $\boldsymbol{U} = \sum_{j=1}^q \boldsymbol{u}_j \boldsymbol{\epsilon}_j^\top$, where $\boldsymbol{\epsilon}_j$ is the jth column of $\boldsymbol{I}$, it easily follows from the basic properties of the vec operation that $\frac{\partial \operatorname{vec}(\boldsymbol{U})}{\partial \operatorname{vec}(\boldsymbol{Y})^\top} = \sum_{j=1}^q (\boldsymbol{\epsilon}_j \otimes \boldsymbol{I}) \frac{\partial \boldsymbol{u}_j}{\partial \operatorname{vec}(\boldsymbol{Y})^\top}$.

As $\boldsymbol{Y}\boldsymbol{u}_j = \lambda_j \boldsymbol{u}_j$ by definition, differentiating both sides over Y_{ik} (i.e. the (i,k)th element of $\boldsymbol{Y}$) and pre-multiplying by $\boldsymbol{u}_j^\top$, we arrive at $\frac{\partial \lambda_j}{\partial Y_{ik}} = \boldsymbol{\epsilon}_i^\top \boldsymbol{u}_j \boldsymbol{u}_j^\top \boldsymbol{\epsilon}_k$, which is in turn equal to the (k,i)th element of $\boldsymbol{u}_j \boldsymbol{u}_j^\top$. Therefore, using matrix notation, $\frac{\partial \lambda_j}{\partial \boldsymbol{Y}} = \boldsymbol{u}_j \boldsymbol{u}_j^\top$, which is equivalent to $\frac{\partial \lambda_j}{\partial \operatorname{vec}(\boldsymbol{Y})} = \operatorname{vec}(\boldsymbol{u}_j \boldsymbol{u}_j^\top)$. Denoting by $\boldsymbol{E}_j$ an $n \times n$ matrix with all its entries equal to zero except for the identity in the (j,j)th entry, $e^{\boldsymbol{\Delta}} = \sum_{j=1}^q \boldsymbol{E}_j e^{\lambda_j}$. Therefore, $\frac{\partial \operatorname{vec}(e^{\boldsymbol{\Delta}})}{\partial \operatorname{vec}(\boldsymbol{Y})^\top} = \sum_{j=1}^q \operatorname{vec}(\boldsymbol{E}_j) e^{\lambda_j} \frac{\partial \lambda_j}{\partial \operatorname{vec}(\boldsymbol{Y})^\top}$. Thus, based on the above, $\frac{\partial \operatorname{vec}(e^{\boldsymbol{Y}})}{\partial \operatorname{vec}(\boldsymbol{Y})^\top}$ can be easily computed.

Finally, using the chain rule along with the definition of the duplication matrix $\boldsymbol{G}_q$, $\operatorname{vec}(\boldsymbol{Y}) = \boldsymbol{G}_q \operatorname{vech}(\boldsymbol{Y})$ Harville (1997), $\frac{\partial \operatorname{vech}(e^{\boldsymbol{Y}})}{\partial \operatorname{vech}(\boldsymbol{Y})^\top}$ can be shown to be equal to $\boldsymbol{H}_q \frac{\partial \operatorname{vec}(e^{\boldsymbol{Y}})}{\partial \operatorname{vec}(\boldsymbol{Y})^\top} \boldsymbol{G}_q$, where $\boldsymbol{H}_q = (\boldsymbol{G}_q^\top \boldsymbol{G}_q)^{-1} \boldsymbol{G}_q^\top$.

To apply the above procedure for the separation strategy prior distribution, we need to first calculate the implied prior distribution for $\boldsymbol{D}$. The separation strategy prior distribution was described in detail in Barnard et al. (2000). However, the normalizing constants of the distributions were not presented, as the focus was mainly on obtaining a posterior sample through MCMC methods. The analytic form of the prior distribution for $\boldsymbol{D}$ is presented below.

4. MISSPECIFYING THE COVARIANCE STRUCTURE IN A LINEAR MIXED MODEL UNDER MAR DROPOUT

To obtain a distribution for a correlation matrix $\boldsymbol{R}$ that has uniform marginal correlations, Barnard et al. (2000) started from a standard Inverse-Wishart distribution, $IW(\boldsymbol{I}_q, \nu)$, i.e.

$$f(\boldsymbol{D}|\nu) = \frac{|\boldsymbol{I}_q|^{\frac{\nu}{2}}}{2^{\frac{\nu q}{2}}\Gamma_q(\frac{\nu}{2})}|\boldsymbol{D}|^{-\frac{\nu+q+1}{2}}\exp\left\{-\frac{1}{2}\text{tr}(\boldsymbol{D}^{-1})\right\},$$

where $\Gamma_q(x) = \pi^{\frac{1}{2}\binom{q}{2}}\prod_{i=1}^{q}\Gamma\{x+(1-i)/2\}$ is the multivariate generalization of the Gamma function. Letting $c_1 = 1/\left\{2^{\frac{\nu q}{2}}\Gamma_q(\frac{\nu}{2})\right\}$ be the normalizing constant,

$$f(\boldsymbol{D}|\nu) = c_1|\boldsymbol{D}|^{-\frac{\nu+q+1}{2}}\exp\left\{-\frac{1}{2}\text{tr}(\boldsymbol{D}^{-1})\right\}.$$

As mentioned in Barnard et al. (2000), the Jacobian of $\boldsymbol{D} \to (\boldsymbol{S}, \boldsymbol{R})$ is given by $2^q\prod_{i=1}^{q}s_i^q$. Thus, the density function of the distribution of $(\boldsymbol{S}, \boldsymbol{R})$ is equal to

$$\begin{aligned}
f(\boldsymbol{S}, \boldsymbol{R}|\nu) &= c_1 2^q|\text{diag}(\boldsymbol{S})\boldsymbol{R}\text{diag}(\boldsymbol{S})|^{-\frac{\nu+q+1}{2}}\prod_{i=1}^{q}s_i^q\exp\left[-\frac{1}{2}\text{tr}\left\{\text{diag}(1/\boldsymbol{S})\boldsymbol{R}^{-1}\text{diag}(1/\boldsymbol{S})\right\}\right] \\
&= c_1 2^q|\boldsymbol{R}|^{-\frac{\nu+q+1}{2}}|\text{diag}(\boldsymbol{S})|^{-(\nu+q+1)}\prod_{i=1}^{q}s_i^q\exp\left\{-\frac{r^{ii}}{2s_i^2}\right\} \\
&= c_1 2^q|\boldsymbol{R}|^{-\frac{\nu+q+1}{2}}\prod_{i=1}^{q}s_i^{-(\nu+1)}\exp\left\{-\frac{r^{ii}}{2s_i^2}\right\},
\end{aligned}$$

where r^{ii} is the (i,i)th element of $\boldsymbol{R}^{-1}$. Then the marginal distribution of $\boldsymbol{R}$ is given by

$$\begin{aligned}
f(\boldsymbol{R}|\nu) &= \int f(\boldsymbol{S}, \boldsymbol{R}|\nu)d\boldsymbol{S} \\
&= c_1 2^q|\boldsymbol{R}|^{-\frac{\nu+q+1}{2}}\prod_{i=1}^{q}\int_0^\infty s_i^{-(\nu+1)}\exp\left\{-\frac{r^{ii}}{2s_i^2}\right\}ds_i.
\end{aligned}$$

To evaluate $\int_0^\infty s_i^{-(\nu+1)}\exp\left\{-\frac{r^{ii}}{2s_i^2}\right\}ds_i$ we make the change of variables $\xi = \frac{r^{ii}}{2s_i^2}$, implying that $ds_i = -\frac{{r^{ii}}^{1/2}\xi^{-3/2}}{2^{3/2}}d\xi$. After some algebra, the integral $\int_0^\infty s_i^{-(\nu+1)}\exp\left\{-\frac{r^{ii}}{2s_i^2}\right\}ds_i$ can be shown to be equal to $r^{ii-\frac{\nu}{2}}2^{\frac{\nu}{2}-1}\int_0^\infty \xi^{\frac{\nu}{2}-1}\exp(-\xi)d\xi$, which is in turn equal to $r^{ii-\frac{\nu}{2}}2^{\frac{\nu}{2}-1}\Gamma(\frac{\nu}{2})$, by the definition of the Gamma function. Thus, it can be easily shown that the marginal distribution of $\boldsymbol{R}$ is equal to

$$f(\boldsymbol{R}|\nu) = \frac{\Gamma(\frac{\nu}{2})^q}{\Gamma_q(\frac{\nu}{2})}|\boldsymbol{R}|^{-\frac{\nu+q+1}{2}}\prod_{i=1}^{q}r^{ii-\frac{\nu}{2}}.$$

Barnard et al. (2000) showed that setting the degrees of freedom equal to $q+1$ (i.e. $\nu = q+1$) results in a uniform marginal distribution for r_{ij}, the (i,j)th element of $\boldsymbol{R}$. This is the approach ($\nu = q+1$) that we adopt when using the separation strategy prior.

We also specified independent prior distributions for the variances $\boldsymbol{S}^2$, which, together with the correlation matrix $\boldsymbol{R}$, completely specify the covariance matrix $\boldsymbol{D}$. However, to get a proper distribution for $\boldsymbol{D}$, we need to also calculate the Jacobian of the transformation of $(\boldsymbol{S}^2, \boldsymbol{R})$ to $\boldsymbol{D}$. For simplicity we present results for $q = 3$. Assume that

$$\boldsymbol{D} = \begin{pmatrix} D_{11} & D_{21} & D_{31} \\ D_{21} & D_{22} & D_{32} \\ D_{31} & D_{32} & D_{33} \end{pmatrix}.$$

Let $\boldsymbol{x}^\top = (D_{11}, D_{22}, D_{33}, r_{21}, r_{31}, r_{32})$ and $\boldsymbol{y}^\top = (D_{11}, D_{22}, D_{33}, D_{21}, D_{31}, D_{32})$, thus $\boldsymbol{x}$ can be written as $\boldsymbol{x}^\top = h(\boldsymbol{y}) = (D_{11}, D_{22}, D_{33}, \frac{D_{21}}{\sqrt{D_{22}D_{11}}}, \frac{D_{31}}{\sqrt{D_{33}D_{11}}}, \frac{D_{32}}{\sqrt{D_{33}D_{22}}})$, which implies that

$$\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{y}^\top} = \begin{pmatrix} \frac{\partial D_{11}}{\partial D_{11}} & \frac{\partial D_{11}}{\partial D_{22}} & \frac{\partial D_{11}}{\partial D_{33}} & \frac{\partial D_{11}}{\partial D_{21}} & \frac{\partial D_{11}}{\partial D_{31}} & \frac{\partial D_{11}}{\partial D_{32}} \\ \frac{\partial D_{22}}{\partial D_{11}} & \frac{\partial D_{22}}{\partial D_{22}} & \frac{\partial D_{22}}{\partial D_{33}} & \frac{\partial D_{22}}{\partial D_{21}} & \frac{\partial D_{22}}{\partial D_{31}} & \frac{\partial D_{22}}{\partial D_{32}} \\ \frac{\partial D_{33}}{\partial D_{11}} & \frac{\partial D_{33}}{\partial D_{22}} & \frac{\partial D_{33}}{\partial D_{33}} & \frac{\partial D_{33}}{\partial D_{21}} & \frac{\partial D_{33}}{\partial D_{31}} & \frac{\partial D_{33}}{\partial D_{32}} \\ \frac{\partial D_{21}/\sqrt{D_{22}D_{11}}}{\partial D_{11}} & \frac{\partial D_{21}/\sqrt{D_{22}D_{11}}}{\partial D_{22}} & \frac{\partial D_{21}/\sqrt{D_{22}D_{11}}}{\partial D_{33}} & \frac{\partial D_{21}/\sqrt{D_{22}D_{11}}}{\partial D_{21}} & \frac{\partial D_{21}/\sqrt{D_{22}D_{11}}}{\partial D_{31}} & \frac{\partial D_{21}/\sqrt{D_{22}D_{11}}}{\partial D_{32}} \\ \frac{\partial D_{31}/\sqrt{D_{33}D_{11}}}{\partial D_{11}} & \frac{\partial D_{31}/\sqrt{D_{33}D_{11}}}{\partial D_{22}} & \frac{\partial D_{31}/\sqrt{D_{33}D_{11}}}{\partial D_{33}} & \frac{\partial D_{31}/\sqrt{D_{33}D_{11}}}{\partial D_{21}} & \frac{\partial D_{31}/\sqrt{D_{33}D_{11}}}{\partial D_{31}} & \frac{\partial D_{31}/\sqrt{D_{33}D_{11}}}{\partial D_{32}} \\ \frac{\partial D_{32}/\sqrt{D_{33}D_{22}}}{\partial D_{11}} & \frac{\partial D_{32}/\sqrt{D_{33}D_{22}}}{\partial D_{22}} & \frac{\partial D_{32}/\sqrt{D_{33}D_{22}}}{\partial D_{33}} & \frac{\partial D_{32}/\sqrt{D_{33}D_{22}}}{\partial D_{21}} & \frac{\partial D_{32}/\sqrt{D_{33}D_{22}}}{\partial D_{31}} & \frac{\partial D_{32}/\sqrt{D_{33}D_{22}}}{\partial D_{32}} \end{pmatrix}$$

$$= \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \hline & & & \frac{1}{\sqrt{D_{22}D_{11}}} & 0 & 0 \\ & & & 0 & \frac{1}{\sqrt{D_{33}D_{11}}} & 0 \\ & & & 0 & 0 & \frac{1}{\sqrt{D_{33}D_{22}}} \end{array}\right)$$

Expanding the determinant of $\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{y}^\top}$ in terms of the cofactors, it follows that $\det\left(\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{y}^\top}\right)$

is equal to the determinant of

$$\begin{pmatrix} \frac{1}{\sqrt{D_{22}D_{11}}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{D_{33}D_{11}}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{D_{33}D_{22}}} \end{pmatrix}, \tag{4.17}$$

which is equal to $\prod_{i=1}^{3} D_{ii}^{-2/2}$. By using mathematical induction, it can be shown that in general the Jacobian is equal to $\prod_{i=1}^{q} D_{ii}^{-(q-1)/2}$.

4.6 Simulation study

To evaluate the performance of the proposed method in correctly identifying the true model, relatively to the performance of the AIC and BIC criteria, we conducted a simulation study. Three simulation scenarios were considered for the number of subjects ($N = 50, 100, 200$); for each scenario we simulated 500 datasets from the models M1 and M2, using the parameter values of the data generating mechanisms given in Subsection 4.2.1. Regarding the parameters of the dropout mechanism, we set $c_1 = 1.714$ and $c_2 = -0.174$, with $Y^\star = 0$.

For both models, the prior distributions for $\boldsymbol{\beta}$ and ω were the $N(\mathbf{0}, 100\boldsymbol{I}_2)$ and Gamma$(0.001, 0.001)$ distributions, respectively. The prior distribution for κ was the $N(0.5, 15)$ distribution truncated to zero, whereas the prior distribution for H was the Beta$(1.5, 1.5)$ distribution. For $\boldsymbol{D}$, we considered both the IW and the separation strategy prior distributions. For the IW distribution, we set ν equal to $q + 2$, which ensures a finite expected value. In M1, we set $\boldsymbol{A} = 4 \times \text{diag}(5, 0.15)$ with $\text{vech}(\boldsymbol{D}_{01}) = (6.46, -0.55, 0.20)^\top$ being the corresponding true value, whereas in M2 we took $\boldsymbol{A} = 5 \times \text{diag}(4, 0.40, 0.10)$ with $\text{vech}(\boldsymbol{D}_{02}) = (4.65, 0.19, 0.19, 0.57, 0.11, 0.14)^\top$ being the corresponding true value, where vech is the "vector-half" operation stacking the columns of the lower triangular part of a symmetric matrix. In the separation strategy prior, for both M1 and M2, we assumed a prior distribution for $\boldsymbol{R}$ resulting in marginally uniformly distributed pairwise correlations (Barnard et al., 2000). In M1, we assumed independent truncated Normal distributions for the diagonal elements

of $\boldsymbol{D}$ with means 5 and 0.15, and associated variances 20 and 0.15, respectively. For the model M2, we also assumed independent truncated Normal distributions for the diagonal elements of $\boldsymbol{D}$ with means 4, 0.40, and 0.10, along with associated variances 20, 0.50, and 0.06, respectively. Equation (4.12) was approximated by the adaptive quadrature rules with 5 points for M1, and by quasi-Monte Carlo integration combined with importance sampling using 10000 draws for M2.

For the penalty term of the BIC criterion, we considered the following approaches: (i) the total number of observations, (ii) the total number of subjects and (iii) the "effective sample size" as proposed by Jones (2011) (i.e. the sum of the inverse of the correlation matrix corresponding to $\boldsymbol{V}_{i,(m_i)}$, summed over subjects). For each simulated dataset, we recorded the preferred model by all the examined model comparison criteria. The results are presented in Table 4.1. For the M1 model, non-convergence of the algorithm used to calculate the posterior mode occurred 1.6% and 0.8% of the time for $N = 50$ and $N = 100$, respectively. As expected, the performance of all examined criteria improved as the sample size increased. When M1 was the true data generating model, all versions of the BIC criterion performed better than the proposed method and the AIC criterion; the proposed method using the separation strategy approach slightly outperformed the IW approach (for N=50, the true model was correctly identified 94.7% and 92.9% of the time, respectively), whereas the AIC criterion had slightly worse performance (for N=50, the true model was correctly identified 90.6% of the time). On the other hand, under M2, all versions of the BIC criterion performed poorly compared to the proposed method and the AIC criterion, especially the one using the total sample size in the definition of the penalty term. The proposed method yielded better results when using the IW rather than the separation strategy approach, especially in small samples (for N=50, the true model was correctly identified 91.0% and 84.0% of the time, respectively). Thus, overall, the IW prior distribution seems to work better than the separation strategy approach in correctly identifying the "true" model in our simulation set-up.

Table 4.1: Proportion of time the true model was selected according to each criterion. 500 replications with N individuals per dataset.

True Model: LMM with random slopes + BM (M1)							True Model: LMM with natural splines (M2)					
N	**PMP(IW)**	**PMP(SEP)**	**AIC**	**BIC1**	**BIC2**	**BIC3**	**PMP(IW)**	**PMP(SEP)**	**AIC**	**BIC1**	**BIC2**	**BIC3**
50	92.9	94.7	90.6	97.2	95.2	96.0	91.0	84.0	87.4	60.8	77.6	74.6
100	98.0	99.2	97.2	99.8	99.0	99.2	97.4	94.8	96.8	84.8	92.2	91.6
200	99.8	100.0	99.6	100.0	100.0	100.0	99.6	99.2	99.6	97.8	98.8	98.8

Abbreviations: LMM, linear mixed model; BM Brownian motion; PMP(IW), criterion based on posterior model probability using the Inverse-Wishart prior; PMP(SEP), criterion based on posterior model probability using the separation strategy prior; BIC1, using total sample size; BIC2, using number of subjects; BIC3, using "effective sample size" Jones (2011).

To evaluate the sensitivity of our proposed method to the choice of prior distributions, we recalculated the PMPs using different prior hyperparameters. Specifically, for the IW approach, we multiplied and also divided the scale matrix $\boldsymbol{A}$ by 1.5, whereas for the separation strategy approach, we multiplied and divided the prior means of the variances by 1.5. As an additional scenario, we set the degrees of freedom of the IW distribution equal to $q+1$, and we increased the prior variances in the separation strategy approach by 50%. The results are presented in Table 4.2, showing that when the sample size is small (N=50), the IW approach is slightly more sensitive to the choice of hyperparameters than the separation-strategy one.

4.7 Application

To illustrate our methodology, we used CD4 data from the CASCADE study up to 5 years since HIV seroconversion, restricting to Europeans (according to the country of birth) infected through sex between men after 1/1/2004. As suggested by many researchers, we used the square root transformation of the CD4 counts, which leads to an almost linear average within-patient decline over time since seroconversion and to an approximate Normal distribution. In total, 6,984 patients met our inclusion criteria; 4,374 (62.6%) were truncated due to ART initiation, 97 (1.4%) due to AIDS onset or death, 660 (9.5%) were considered as lost to follow up, whereas the remaining 1853

Table 4.2: Proportion of time the true model was selected according to each criterion. 500 replications with N individuals per dataset. Three sensitivity analysis scenarios regarding the choice of prior hyper-parameters were performed. **Sensitivity analysis scenario (I)**: The scaled matrix $\boldsymbol{A}$ in the Inverse-Wishart approach and the prior means for the variances in the separation strategy approach were increased by 50%, **Sensitivity analysis scenario (II)**: The scaled matrix $\boldsymbol{A}$ in the Inverse-Wishart approach and the prior means for the variances in the separation strategy approach were decreased by 50%, and **Sensitivity analysis scenario (III)**: The degrees of freedom ν in the Inverse-Wishart approach were set to $q+1$, whereas the prior variances for the variances in the separation strategy approach were increased by 50%.

True Model: model M1			**True Model: model M2**	
Sensitivity analysis scenario (I)				
$\boldsymbol{N}$	**PMP(IW)**	**PMP(SEP)**	**PMP(IW)**	**PMP(SEP)**
50	95.7	95.9	85.8	82.2
100	98.6	99.4	96.8	94.2
200	100.0	100.0	99.6	99.2
Sensitivity analysis scenario (II)				
50	91.5	94.5	91.6	86.1
100	97.8	99.0	97.2	95.2
200	99.8	100.0	99.6	99.4
Sensitivity analysis scenario (III)				
50	94.7	95.7	88.4	81.6
100	98.2	99.2	97.0	94.4
200	100.0	100.0	99.6	99.2

Abbreviations: Model M1, linear mixed model with random slopes + fractional BM;
Model M2, linear mixed model with natural splines for the random effects;
PMP(IW), criterion based on posterior model probability using the Inverse-Wishart prior;
PMP(SEP), criterion based on posterior model probability using the separation strategy prior.

(26.5%) patients were administratively censored at the end of their follow-up time. As the proportion of potentially informative dropouts (death or AIDS onset) is practically negligible, throughout this chapter we assume that the mechanisms leading to incomplete data are MAR. Visual inspection of the data also reveals great heterogeneity of the individual trajectories over time as well as highly imbalanced frequency of clinic visits, with the median number of observations per individual being 4, the corresponding number ranging from 1 to 27 and the IQR being 2-7.

We applied the models M1, M2, and M3, along with an LMM using a non-fractional BM process on top of random intercepts and slopes (M4). In all models, we assumed a linear population CD4 decline on the square root scale and we also used age at seroconversion, in four groups, [15-25), [25-35), [35,45) and [45,) years, as a covariate affecting both the CD4 counts at HIV seroconversion (i.e. baseline) and the CD4 decline.

We applied our proposed method based on the IW and the separation strategy approaches, the AIC criterion, and the three versions of the BIC criterion to identify the most plausible model. The prior distributions assigned to $\boldsymbol{\beta}$ were independent Normal distributions with variances equal to 100. The prior means for the population intercept and slope were set to 23 $\sqrt{\text{cells}/\mu L}$ and -1.30 $\sqrt{\text{cells}/\mu L}$/year, respectively, whereas the prior means for the effects of age on the baseline CD4 values and the CD4 slopes were set to zero; for ω, we assumed the Gamma$(0.001, 0.001)$ distribution. The results are presented in Table 4.3. The preferred model according to all criteria was the model M1; among all the examined models, the PMPs of M1 were 99.998% and 99.656% based on the IW and the separation prior approach, respectively. Also, the likelihood ratio test comparing the fractional BM model to the non-fractional one yielded a strongly significant result ($p<0.001$). The effects of age on the baseline CD4 values were of similar magnitude in all models. All models also showed that the CD4 decline tends to accelerate at higher ages. However, the CD4 slope estimate in the baseline group, i.e. [15-25) years, was quite steeper in M3 (-0.94) compared to the corresponding one in M1 (-0.74) and in M2 (-0.85). Moreover, the differences between the slope estimates from

the M3 model and the corresponding estimates from the remaining models were more pronounced at higher ages (Table 4.3), where further analyses revealed higher dropout rate.

4.8 Discussion

In this chapter, we evaluated the performance of specific LMMs in estimating trends of longitudinal markers under misspecified covariance structure and specific MAR dropout mechanisms. We analytically calculated the asymptotic bias in the estimated marker trends of LMMs using covariance structures that have been used in real data applications. The results showed that assuming a simple "random intercept and slope" structure when the true one is substantially more complex can lead to serious biases, especially when dropout is intense (e.g. starting ART at higher CD4 counts, thus earlier on during the course of the disease) and the average marker profiles are more elaborate than a linear slope. We also compared the approach of using a fractional BM process on top of a random intercept and slope structure with the approach of using natural splines for the random effects in terms of their robustness to misspecified covariance structure. Assuming that the former is the true data generating model, we calculated the bias in the fixed-effect estimates when the latter is fitted to data, and vice versa. For the estimation of a single slope, the two approaches performed almost equally well, though when the "true" population average marker evolution was based on piece-wise linear or natural splines, the fractional BM model yielded less biased estimates. We also considered a two groups scenario; the bias in the estimated slope difference increased with higher MAR dropout probabilities, as well as when the dropout mechanisms were substantially different in the two groups. We also proved analytically that if the true slope difference is zero and the dropout mechanism is the same in the two groups, the estimate of the slope difference from the misspecified model is always unbiased. Although we focused on discrete dropout mechanisms, additional results suggested similar degree of bias when the dropout times are continuous.

Table 4.3: Modeling CD4 cell count trends in the CASCADE data. Results from linear mixed models with different covariance structures. The most plausible model is identified by the examined model comparison criteria.

Parameter	**M3**	**M2**	**M4**	**M1**
Intercept $\sqrt{\text{cells}/\mu L}$				
Age:[15,25)	23.48 (23.14,23.81)	23.35 (23.02,23.69)	23.28 (22.94,23.61)	23.27 (22.94,23.61)
Age:[25,35)/[15,25)	0.32 (-0.07,0.71)	0.34 (-0.05,0.73)	0.35 (-0.04,0.74)	0.35 (-0.04,0.74)
Age:[35,45)/[15,25)	0.04 (-0.37,0.45)	0.03 (-0.38,0.44)	0.02 (-0.38,0.43)	0.02 (-0.39,0.42)
Age:[45,)/[15,25)	-0.29 (-0.79,0.21)	-0.32 (-0.82,0.18)	-0.30 (-0.8,0.2)	-0.31 (-0.81,0.19)
Rate of change $\sqrt{\text{cells}/\mu L}$/year				
Age:[15,25)	-0.94 (-1.09,-0.78)	-0.82 (-0.97,-0.68)	-0.75 (-0.89,-0.6)	-0.74 (-0.89,-0.59)
Age:[25,35)/[15,25)	-0.34 (-0.52,-0.16)	-0.34 (-0.52,-0.17)	-0.35 (-0.52,-0.18)	-0.35 (-0.52,-0.18)
Age:[35,45)/[15,25)	-0.43 (-0.62,-0.24)	-0.37 (-0.55,-0.19)	-0.38 (-0.56,-0.2)	-0.38 (-0.55,-0.2)
Age:[45,)/[15,25)	-0.51 (-0.76,-0.27)	-0.45 (-0.69,-0.22)	-0.45 (-0.69,-0.22)	-0.45 (-0.68,-0.22)
Kappa			5.06	5.46
Hurst				0.36
PMP(IW) (%)	<0.001	<0.001	0.002	99.998
PMP(SEP) (%)	<0.001	<0.001	0.344	99.656
AIC	173988.21	173513.98	173234.7	173224.85
BIC-1	174089.34	173640.39	173344.26	173342.84
BIC-2	174070.42	173616.75	173323.76	173320.77
BIC-3	174072.38	173619.26	173325.86	173323.01

For the population intercept and rate of change estimates, the [15,25) age group is the baseline group, with the remaining estimates denoting the differences compared to the estimates for the baseline age group.

Abbreviations: BIC1, using total sample size; BIC2, using number of subjects; BIC3, using "effective sample size" Jones (2011); PMP(IW), posterior model probability using the Inverse-Wishart prior distribution; PMP(SEP): posterior model probability using separation strategy prior distribution;

M3, random intercepts and slopes; PMP(IW): $\boldsymbol{D} \sim IW\{\text{diag}(5, 0.5), 4\}$; PMP(SEP): $\text{diag}(\boldsymbol{D}) \sim N\{(5, 0.5), \text{diag}(20, 0.15)\}$ truncated on $(0, \infty)$.

M2, natural splines on the random effects; PMP(IW): $\boldsymbol{D} \sim IW\{\text{diag}(4, 0.40, 0.10), 5\}$; PMP(SEP): $\text{diag}(\boldsymbol{D}) \sim N\{(4, 0.40, 0.10), \text{diag}(20, 0.50, 0.06)\}$ truncated on $(0, \infty)$.

M4, Brownian motion + random intercepts and slopes; $\kappa \sim N(0.5, 15)$ truncated on $(0, \infty)$; PMP(IW): $\boldsymbol{D} \sim IW\{\text{diag}(5, 0.15), 4\}$; PMP(SEP): $\text{diag}(\boldsymbol{D}) \sim N\{(5, 0.15), \text{diag}(20, 0.15)\}$ truncated on $(0, \infty)$.

M1, fractional Brownian motion + random intercepts and slopes; $H \sim \text{Beta}(1.5, 1.5)$ and $\kappa \sim N(0.5, 15)$ truncated on $(0, \infty)$; PMP(IW): $\boldsymbol{D} \sim IW\{\text{diag}(5, 0.15), 4\}$; PMP(SEP): $\text{diag}(\boldsymbol{D}) \sim N\{(5, 0.15), \text{diag}(20, 0.15)\}$ truncated on $(0, \infty)$.

We have adopted Bayesian model comparison based on the posterior model probabilities to distinguish between the approach of using splines in the design matrix of the random effects and the approach of adding a BM process to a random intercept and slope structure in real data applications. To our knowledge, this is the first work considering Bayesian model comparison based on the posterior model probabilities of LMMs with/without additional BM processes that can have any, and potentially different, random-effects specifications. Similarly to the method applied to obtain the restricted maximum likelihood estimates, the fixed effects were explicitly integrated out, resulting in a usually low dimensional integral over the variance components. Our approach, though, relies on numerical integration methods based either on the adaptive quadrature method or on quasi Monte Carlo along with importance sampling to approximate the integrals over the variance components. Since the integrand is centered at the posterior mode and scaled based on the inverse curvature at the posterior mode, both approaches generalize the Laplace method in some sense. A limitation of the adaptive quadrature is that the number of evaluations grows exponentially as the number of dimensions increases, in which cases, quasi Monte Carlo using a prespecified number of draws can be used instead, though it might need a relatively large number of draws to achieve the desired accuracy. Two popular choices for the prior distribution of the covariance matrix of the random effects were considered: (i) the Inverse-Wishart distribution and (ii) the separation strategy approach. The performance of the proposed method was compared to the AIC and BIC criteria through simulation studies. Overall, the Inverse-Wishart approach worked better than the separation strategy one, though the former seemed to be slightly more sensitive to the choice of prior hyperparameters in small samples. The AIC criterion had better overall performance than the BIC criterion, with the version of BIC using the total sample size having by far the worst performance, especially in small samples. Similar findings regarding the AIC and BIC criteria were observed in Keselman et al. (1998).

The examined models were fitted to recent data from the CASCADE study, with treatment initiation being the dominant reason for dropout. The preferred model by all

criteria was the one including a fractional BM process on top of random intercept and slope. The slope estimate in the simple random intercept and slope model was quite steeper compared to that from the fractional BM model, with the difference being more pronounced at higher ages where the dropout rate was higher. Given that the fractional BM model was shown to be the most plausible one, the results confirmed our analytical findings suggesting biased slope estimates when the fitted covariance structure is simpler than the true one. Thus, methods that are based on results from simple random intercept and slope models fitted to CASCADE CD4 data (Pantazis et al., 2019b) could be reevaluated applying either LMMs including also a BM process or LMMs with splines for the random effects.

In this chapter, we compared, in terms of bias in the fixed-effect estimates, some frequently used approaches for modeling the covariance structure. We assumed specific MAR dropout mechanisms with the parameters of the "true" data generating mechanisms based on our motivating example. However, our approach can be very easily modified to examine any covariance structure of LMMs with/without an additional BM process, other types of MAR dropout, and different data generating settings (e.g. different "true" parameter values and frequency of visits). The effects on other aspects of inference, though, such as the type I error of Wald tests or the mean squared error, have not been addressed in this paper. Moreover, in this project, we have considered only fixed observation/visiting times, for an interesting exploration of MAR in continuous time, we refer to Farewell et al. (2018). Another source of bias not considered in this paper occurs when the random effects deviate from the normal distribution. Nevertheless, in random slope models, it has been shown that the fixed-effect estimates turn out to be quite robust to violations of the normality assumption (Song et al., 2002). The effect of misspecified covariance structure could also be investigated in a joint modeling framework assuming MNAR missingness. However, as fitting joint models is not trivial, including additional stochastic processes might lead to convergence problems.

To conclude, we have shown that assuming an over-simplistic covariance structure can lead to seriously biased fixed-effect estimates in LMMs, with the magnitude of the

bias increasing with increasing probability of MAR dropout. To reduce the induced bias in the fixed-effect estimates of an LMM due to potentially misspecified covariance structure, the approach of adding a BM process on top of a random intercept and slope structure or the approach of using an LMM with natural splines for the random effects could be adopted as we have shown that they lead to much smaller bias under specific covariance structure misspecifications, though the BM approach had better performance in most cases. To discriminate between different approaches of modeling the covariance structure, our proposed Bayesian model comparison had better overall performance than the AIC or BIC criteria in simulation studies, and thus it could be applied in real data applications, especially when there is prior information on the parameters to be estimated.

Chapter 5

Joint modeling of longitudinal and competing-risk data accounting for failure cause misclassification

5.1 Introduction

In the previous chapter, we investigated the bias of fixed-effect estimates of an LMM when fitted to incomplete data due to MAR dropout under covariance structure misspecification. We showed that assuming a simple random intercept and slope structure when the true one is more complex can lead to serious bias especially when the MAR dropout probabilities are high. Two alternative approaches were formally compared in terms of bias under covariance structure misspecifications: (a) a fractional BM process on top of a random intercept and slope structure, and (b) a more elaborate random-effect specification through natural splines of time. We also proposed a Bayesian model comparison criterion to identify the model with the best fit to the data.

In this chapter, we are concerned with joint modeling of longitudinal marker data

and multiple dropout types (competing risks) when the true failure cause is not observed for all individuals. Recall that, as it is mentioned in Chapter 2, in SREM-based joint models, the mechanism causing termination of data collection is assumed to be MNAR as it depends on the random effects, which are unobserved. In Section 5.2, we present the proposed approach of jointly modeling a continuous disease marker over time and competing risks, where the failure submodels are specified in terms of CIFs parametrized through a general family of transformation models. The extension of the proposed model when there is failure cause misclassification is described in Section 5.3. In Section 5.4, we describe how, solely on the basis of the proposed model, a posterior sample for multistate/transition probabilities of mutually exclusive states defined by unobserved marker data and competing risks can be obtained. We also propose a marginalized deviance information criterion to assess the fit of the models in Section 5.5. In Section 5.6, we carry out a simulation study to evaluate the performance of the proposed approach, whereas, the proposed methodology is applied to real data from the East Africa International Epidemiologic Databases to Evaluate AIDS (IeDEA) study. Finally, Section 5.8 presents concluding remarks and discusses limitations along with possible extensions.

5.2 Proposed model

5.2.1 Marker model

For the longitudinal marker model, we use an LMM model of the form $y_i(t) = \boldsymbol{x}_i^\top(t)\boldsymbol{\beta} + \boldsymbol{z}_i^\top(t)\boldsymbol{b}_i + \epsilon_i(t)$ (Laird and Ware, 1982). Recall that $\boldsymbol{x}_i^\top(t)$ and $\boldsymbol{z}_i^\top(t)$ denote the fixed and random effects design matrices at time t, respectively, $\boldsymbol{\beta}$ and $\boldsymbol{b}_i \sim N(\boldsymbol{0}, \boldsymbol{D})$ denote the fixed and random effects, respectively. Also, $\epsilon_i(t) \sim N(0, \omega^{-1})$ denotes the within-subject residuals with ω being the within-subject precision. As usually assumed in the joint modelling literature (Rizopoulos, 2012b), the "true" marker value at time t is defined as $m_i(t) = \boldsymbol{x}_i^\top(t)\boldsymbol{\beta} + \boldsymbol{z}_i^\top(t)\boldsymbol{b}_i$ and the history of "true" values up to time t are denoted by $M_i(t) = \{m_i(s) : 0 \leq s \leq t\}$. The vector of the marker measurements intended

to be collected on the ith subject is denoted by $\boldsymbol{Y}_i^\top = \{Y_i(t_{i1}), Y_i(t_{i2}), \ldots, Y_i(t_{iQ_i})\}$, where $t_{i1}, \ldots t_{iQ_i}$ are the observation times and Q_i is the number of marker measurements on subject i. The corresponding design matrices for the fixed and random effects at times $t_{i1}, \ldots t_{iQ_i}$ are denoted by $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ respectively. The parameters relating to the marker model are denoted by $\boldsymbol{\theta}_L^\top = (\boldsymbol{\beta}^\top, \text{vech}(\boldsymbol{D})^\top, \omega)$ and $\boldsymbol{b}^\top = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_N)$.

5.2.2 Competing-risk survival models

Let $T_i^\star$ be the time to the first event for the ith individual and $K_i \in \{1, 2, \ldots, K\}$ be the corresponding failure cause. We propose to model the CIFs for all causes simultaneously conditional on the "true" marker values, i.e.

$$F_{ik}\{t|M_i(t), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_{tk}\} = \Pr\{T_i^\star \leq t, K_i = k|M_i(t), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_{tk}\}, \tag{5.1}$$

where $\boldsymbol{w}_{ik}$ is a vector of baseline covariates for cause k and individual i and $\boldsymbol{\theta}_{tk}$ is the parameter vector for cause k. Note that (5.1) depends on the longitudinal model parameters ($\boldsymbol{\beta}$ and $\boldsymbol{b}_i$) through the history of the true marker values, $M_i(t)$. Since all CIFs are modelled simultaneously, the all-cause CIFs should be bounded by 1 at each failure time. To formally account for these boundness constraints, we assume that all CIFs increase over time up to a common point τ_i at which the all-cause CIF approaches 1 and they become plateau thereafter, i.e.

$$\begin{aligned} F_{ik}\{t|M_i(t), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_{tk}\} &= F_{ik}^M\{t|M_i(t), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_{tk}\} I(0 \leq t < \tau_i) \\ &+ F_{ik}^M\{\tau_i|M_i(\tau_i), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_{tk}\} I(t \geq \tau_i), \end{aligned} \tag{5.2}$$

where $F_{ik}^M\{t|M_i(t), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_{tk}\}$ is a certain parametric model for the CIF of cause k conditional on the true marker values, $M_i(t)$, and some baseline covariates, $\boldsymbol{w}_{ik}$. To be more formal, τ_i depends on the values of $(\boldsymbol{\beta}^\top, \boldsymbol{\theta}_t^\top, \boldsymbol{b}_i^\top)$, i.e. $\tau_i \equiv \tau_i(\boldsymbol{\beta}, \boldsymbol{\theta}_t, \boldsymbol{b}_i) \equiv \sup_t \left[t : \sum_{k=1}^K F_{ik}^M\{t|M_i(t), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_{tk}\} < 1\right]$, where $\boldsymbol{\theta}_t = (\boldsymbol{\theta}_{t1}^\top, \boldsymbol{\theta}_{t2}^\top, \ldots, \boldsymbol{\theta}_{tK}^\top)^\top$. In other words, τ_i is the upper limit of the support of the distribution of the survival time $T_i^\star$, and if $\sum_{k=1}^K F_{ik}^M\{t|M_i(t), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_{tk}\} < 1$ for any $t > 0$, $\tau_i = \infty$. The motivation for assuming (5.2) may become clearer when considering the survival likelihood conditionally

on the random effects under non-informative right censoring (Jeong and Fine, 2006), in which case we are only able to observe $T_i = \min(T_i^\star, C_i)$, with C_i being the corresponding censoring time. As a convention, we assume that $K_i = 0$ denotes right censoring. In this case, the survival likelihood is equal to

$$\begin{aligned} f\{T_i, K_i | M_i(T_i), \boldsymbol{w}_i; \boldsymbol{\theta}_t\} &= \prod_{k=1}^{K} f_{ik}\{T_i | M_i(T_i), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_{tk}\}^{\delta_{ik}} \\ &\times S_i\{T_i | M_i(T_i), \boldsymbol{w}_i; \boldsymbol{\theta}_t\}^{1-\delta_i}, \end{aligned} \tag{5.3}$$

where $f_{ik}\{x|M_i(x), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_{tk}\} = I(0 < x < \tau_i)\partial F_{ik}^M\{x|M_i(x), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_{tk}\}/\partial x$ is the density function for cause k, $S_i\{T_i|M_i(T_i), \boldsymbol{w}_i; \boldsymbol{\theta}_t\} = 1 - \sum_{k=1}^{K} F_{ik}\{T_i|M_i(T_i), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_{tk}\}$ is the overall survival function, and $\boldsymbol{w}_i = (\boldsymbol{w}_{i1}^\top, \boldsymbol{w}_{i2}^\top, \ldots, \boldsymbol{w}_{iK}^\top)^\top$.

For some specific set of parameter values, $(\boldsymbol{\beta}, \boldsymbol{\theta}_t, \boldsymbol{b}_i)$, suppose that the assumed model yields an all-cause CIF evaluated at the observed survival time, T_i, greater than or equal to 1, i.e. $\sum_{k=1}^{K} F_{ik}^M\{T_i|M_i(T_i), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_{tk}\} \geq 1$. By the definition of τ_i, it is implied that T_i does not lie within the support of $T_i^\star$ given the parameter values, $(\boldsymbol{\beta}, \boldsymbol{\theta}_t, \boldsymbol{b}_i)$, i.e. $T_i \geq \tau_i(\boldsymbol{\beta}, \boldsymbol{\theta}_{tk}, \boldsymbol{b}_i)$. By (5.2), it is further implied that both $f_{ik}\{T_i|M_i(T_i), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_{tk}\}$ and $S_i\{T_i|M_i(T_i), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_t\}$ are equal to zero, ensuring that the likelihood function is equal to zero. Thus, (5.3) is equivalent to including the model-based CIF, $F_{ik}^M\{T_i|M_i(T_i), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_{tk}\}$, and its derivative along with the indicator function $I[\sum_{k=1}^{K} F_{ik}^M\{T_i|M_i(T_i), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_{tk}\} < 1]$ in (5.3). Therefore, our approach is consistent with the work of Salmerón et al. (2015), who, under a Bayesian paradigm, incorporated a similar indicator function in the likelihood of the log-Binomial regression model. In this work, though, the proposed model itself leads to zero likelihood when the constraints are violated.

We propose to model the CIFs using the class of generalized odds rate transformation models (e.g. Bakoyannis et al., 2017, Dabrowska and Doksum, 1988, Jeong and

Fine, 2006)

$$F_{ik}^{M}\{t|M_i(t),\boldsymbol{w}_{ik};\boldsymbol{\theta}_{tk}\} = 1-\exp\left\{-\int_0^t e^{\boldsymbol{B}_k^{\top}(s)\boldsymbol{\psi}_k+\boldsymbol{\gamma}_k^{\top}\boldsymbol{w}_{ik}+\alpha_k m_i(s)}ds\right\}$$

SREM-CIF-1 (5.4)

$$F_{ik}^{M}\{t|M_i(t),\boldsymbol{w}_{ik};\boldsymbol{\theta}_{tk}\} = 1-\left\{1+c_k\int_0^t e^{\boldsymbol{B}_k^{\top}(s)\boldsymbol{\psi}_k+\boldsymbol{\gamma}_k^{\top}\boldsymbol{w}_{ik}+\alpha_k m_i(s)}ds\right\}^{-1/c_k}$$

SREM-CIF-2 (5.5)

The model SREM-CIF-1 is a proportional subdistribution hazard model (Deslandes and Chevret, 2010, Fine and Gray, 1999) since $\lambda_{ik}^{M}\{t|m_i(t),\boldsymbol{w}_{ik};\boldsymbol{\theta}_{tk}\} = \exp\{\boldsymbol{B}_k^{\top}(t)\boldsymbol{\psi}_k + \boldsymbol{\gamma}_k^{\top}\boldsymbol{w}_{ik} + \alpha_k m_i(t)\}$, where $\lambda_{ik}^{M}\{t|m_i(t),\boldsymbol{w}_{ik};\boldsymbol{\theta}_{tk}\}$ is the assumed subdistribution hazard function. Moreover, the model SREM-CIF-2 is a generalization of SREM-CIF-1, which reduces to SREM-CIF-1 as $c_k \to 0$ (Jeong and Fine, 2006). $\boldsymbol{B}_k^{\top}(s)$ denotes a B-splines basis matrix at time s with associated parameter $\boldsymbol{\psi}_k$ and $\boldsymbol{\gamma}_k$ measuring the effect of the baseline covariates on the kth CIF. The parameters α_k $(k = 1,\ldots,K)$ correspond to the effects of the true marker values on the CIFs and indicate the level of association between the marker and the survival submodels (referred to as the association parameters from now on). Under SREM-CIF-1, $\exp(\alpha_k)$ denotes the relative increase in the kth subdistribution hazard function at time t resulting from one unit increase in $m_i(t)$ at the same time point. The interpretation of the parameters of SREM-CIF-2 does not seem so appealing in general, but assuming $c_k = 1$ a proportional rate of odds increase model is implied, i.e. $\partial(F_{ik}^{M}\{t|M_i(t),\boldsymbol{w}_{ik};\boldsymbol{\theta}_{tk}\}/[1-F_{ik}^{M}\{t|M_i(t),\boldsymbol{w}_{ik};\boldsymbol{\theta}_{tk}\}])/\partial t = \exp\{\boldsymbol{B}_k^{\top}(t)\boldsymbol{\psi}_k + \boldsymbol{\gamma}_k^{\top}\boldsymbol{w}_{ik} + \alpha_k m_i(t)\}$. Throughout this paper, we assume that the parameters c_k are known, as trying to estimate them can lead to non-identifiability issues or may require very large datasets (Zeng et al., 2006).

5.2.3 Bayesian inferential procedures

All marker data measurements after time T_i are not observed, thus, supposing that the occurrence of T_i results in $M_i = m_i$, and thus, $\boldsymbol{Y}_{i,(m_i)} = (Y_{i1},\ldots,Y_{im_i})$ denotes the observed marker values. Based on Equation (2.25), the observed data likelihood of the joint models is equal to $\prod_{i=1}^{N}\int f(\boldsymbol{Y}_{i,(m_i)}|\boldsymbol{b}_i;\boldsymbol{\theta}_L)f(\boldsymbol{b}_i;\boldsymbol{\theta}_L)f\{T_i,K_i|M_i(T_i),\boldsymbol{w}_i;\boldsymbol{\theta}_t\}d\boldsymbol{b}_i$,

which requires multidimensional integration over the random effects. Since such integration, especially given the constrained space due to (5.2), is challenging, we rely on a Bayesian inferential procedure based on a Markov chain Monte Carlo (MCMC) algorithm. Letting $\boldsymbol{\theta} = (\boldsymbol{\theta}_L^\top, \boldsymbol{\theta}_t^\top)^\top$ be the whole parameter vector of the model and $\mathcal{D} = \left\{(\boldsymbol{Y}_{i,(m_i)}, \boldsymbol{X}_{i,(m_i)}, \boldsymbol{Z}_{i,(m_i)}, T_i, K_i, \boldsymbol{w}_i), i = 1, 2, \ldots, N\right\}$ be the observed data, the posterior distribution of all unknown quantities is proportional to

$$f(\boldsymbol{\theta}, \boldsymbol{b}|\mathcal{D}) \propto f(\boldsymbol{\theta}) \prod_{i=1}^{N} \left[f(\boldsymbol{Y}_{i,(m_i)}|\boldsymbol{b}_i; \boldsymbol{\theta}_L) f(\boldsymbol{b}_i; \boldsymbol{\theta}_L) f\{T_i, K_i|M_i(T_i), \boldsymbol{w}_i; \boldsymbol{\theta}_t\} \right], \tag{5.6}$$

where $f(\boldsymbol{\theta})$ is the prior distribution of the parameters. The integrals involved in the definition of the CIFs in (5.4) and (5.5) can be accurately approximated by applying numerical quadrature based on the Gauss–Legendre rule with 30 nodes. A Normal prior distribution, $N(\boldsymbol{\mu}_0, \boldsymbol{C}_0)$, is used for $\boldsymbol{\beta}$, a Gamma(λ_1, λ_2) for ω and a Normal, $N(\boldsymbol{\mu}_0^s, \boldsymbol{C}_0^s)$, distribution for $\boldsymbol{\theta}_t$. For the covariance matrix of the random effects, $\boldsymbol{D}$, we assumed the Inverse-Wishart $IW(\boldsymbol{A}, df)$ distribution, $f(\boldsymbol{D}) \propto |\boldsymbol{D}|^{-(df+q+1)/2} \exp\{-\operatorname{tr}(\boldsymbol{D}^{-1}\boldsymbol{A})/2\}$. Given the likelihood and prior specification, it follows that the conditional posterior distribution of ω is Gamma$\{\frac{n}{2} + \lambda_1, \lambda_2 + \frac{1}{2}\sum_{i=1}^{N}(\boldsymbol{Y}_{i,(m_i)} - \boldsymbol{X}_{i,(m_i)}\boldsymbol{\beta} - \boldsymbol{Z}_{i,(m_i)}\boldsymbol{b}_i)^\top(\boldsymbol{Y}_{i,(m_i)} - \boldsymbol{X}_{i,(m_i)}\boldsymbol{\beta} - \boldsymbol{Z}_{i,(m_i)}\boldsymbol{b}_i)\}$, where $n = \sum_{i=1}^{N} m_i$ is the total sample size. The corresponding conditional posterior for $\boldsymbol{D}$ is $IW(\boldsymbol{A} + \sum_{i=1}^{N} \boldsymbol{b}_i\boldsymbol{b}_i^\top, df + N)$. For the remaining parameters, we used Metropolis-Hasting schemes. For $\boldsymbol{\beta}$, we used the conditional posterior distribution given the marker as the proposal density, i.e. $q(\boldsymbol{\beta}^{can}|\mathcal{D}, \boldsymbol{b}; \omega) \sim N(\boldsymbol{\mu}_1, \boldsymbol{C}_1)$, where $\boldsymbol{C}_1 = (\boldsymbol{C}_0^{-1} + \omega\sum_{i=1}^{N} \boldsymbol{X}_{i,(m_i)}^\top \boldsymbol{X}_{i,(m_i)})^{-1}$ and $\boldsymbol{\mu}_1 = \boldsymbol{C}_1\{\boldsymbol{C}_0^{-1}\boldsymbol{\mu}_0 + \omega\sum_{i=1}^{N} \boldsymbol{X}_{i,(m_i)}^\top(\boldsymbol{Y}_{i,(m_i)} - \boldsymbol{Z}_{i,(m_i)}\boldsymbol{b}_i)\}$. It can be shown that the acceptance probability is equal to $p = \min\left\{1, \prod_{i=1}^{N} f\{T_i, K_i|M_i^{can}(T_i), \boldsymbol{w}_i; \boldsymbol{\theta}_t\} / \prod_{i=1}^{N} f\{T_i, K_i|M_i(T_i), \boldsymbol{w}_i; \boldsymbol{\theta}_t\}\right\}$, where $M_i^{can}(T_i)$ and $M_i(T_i)$ denote the true marker values up to time T_i evaluated at $\boldsymbol{\beta}^{can}$ and $\boldsymbol{\beta}$ (the current MCMC value), respectively. It needs to be emphasized that any proposed value leading to an all-cause model-based CIF greater than 1, i.e. $\sum_{k=1}^{K} F_{ik}^M\{t|M_i(t), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_{tk}\} > 1$, is immediately rejected as the posterior ratio is equal to zero. Thus, calculation of $\tau_i(\boldsymbol{\beta}, \boldsymbol{\theta}_t, \boldsymbol{b}_i)$ is not required within the MCMC algorithm. We now present the conditional posterior distributions of the random effects and the

survival parameters, together with the Metropolis-Hasting algorithms used to update their values within the MCMC algorithm.

5.2.3.1 Conditional posterior distribution of the random effects

The conditional posterior distribution of the random effects is proportional to

$$\begin{aligned} f(\boldsymbol{b}_i|\mathcal{D};\boldsymbol{\theta}) \quad &\propto \quad \exp\left[-\frac{1}{2}\boldsymbol{b}_i^\top(\boldsymbol{D}^{-1}+\omega\boldsymbol{Z}_{i,(m_i)}^\top\boldsymbol{Z}_{i,(m_i)})\boldsymbol{b}_i+\omega\boldsymbol{b}_i^\top\boldsymbol{Z}_{i,(m_i)}^\top(\boldsymbol{y}_{i,(m_i)}-\boldsymbol{X}_{i,(m_i)}\boldsymbol{\beta})\right] \\ &\times \quad \prod_{k=1}^{K} f_{ik}^M\{T_i|M_i(T_i),\boldsymbol{w}_{ik};\boldsymbol{\theta}_{tk}\}^{\delta_{ik}}\left[1-\sum_{k=1}^{K}F_{ik}^M\{T_i|M_i(T_i),\boldsymbol{w}_{ik};\boldsymbol{\theta}_{tk}\}\right]^{1-\delta_i} \\ &\times \quad I\left[\sum_{k=1}^{K}F_{ik}^M\{T_i|M_i(T_i),\boldsymbol{w}_{ik};\boldsymbol{\theta}_{tk}\}<1\right], \end{aligned} \tag{5.7}$$

where $f_{ik}^M\{x|M_i(x),\boldsymbol{w}_{ik};\boldsymbol{\theta}_{tk}\}=\partial F_{ik}^M\{x|M_i(x),\boldsymbol{w}_{ik};\boldsymbol{\theta}_{tk}\}\partial x$. To update the value of $\boldsymbol{b}_i$ in each cycle of the MCMC algorithm, starting from the posterior mode using only the marker model, i.e. $\boldsymbol{\mu}_{\boldsymbol{b}_i}=(\boldsymbol{D}^{-1}+\omega\boldsymbol{Z}_{i,(m_i)}^\top\boldsymbol{Z}_{i,(m_i)})^{-1}\omega\boldsymbol{Z}_{i,(m_i)}^\top(\boldsymbol{y}_{i,(m_i)}-\boldsymbol{X}_{i,(m_i)}\boldsymbol{\beta})$, we carry out a single Newton Raphson step to maximize (5.7), i.e. $\boldsymbol{b}_i^\star=\boldsymbol{\mu}_{\boldsymbol{b}_i}+\mathcal{I}(\boldsymbol{\mu}_{\boldsymbol{b}_i})^{-1}\mathcal{U}(\boldsymbol{\mu}_{\boldsymbol{b}_i})$, where $\mathcal{U}(\boldsymbol{b}_i)=\frac{\partial\log f(\boldsymbol{b}_i|\mathcal{D};\boldsymbol{\theta})}{\partial\boldsymbol{b}_i}$ and $\mathcal{I}(\boldsymbol{b}_i)=-\frac{\partial^2\log f(\boldsymbol{b}_i|\mathcal{D};\boldsymbol{\theta})}{\partial\boldsymbol{b}_i\partial\boldsymbol{b}_i^\top}$. However, in very rare cases, $\boldsymbol{\mu}_{\boldsymbol{b}_i}$ might not fulfill the boundness constraint, i.e. it might be the case that $f(\boldsymbol{\mu}_{\boldsymbol{b}_i}|\mathcal{D};\boldsymbol{\theta})=0$. To avoid problems with the logarithm of a function that is equal to zero, we calculated $\mathcal{U}(\boldsymbol{b}_i)$ and $\mathcal{I}(\boldsymbol{b}_i)$ ignoring the boundness constraint and evaluated $\frac{\partial\log\left[1-\sum_{k=1}^{K}F_{ik}^M\{T_i|M_i(T_i),\boldsymbol{w}_{ik};\boldsymbol{\theta}_{tk}\}\right]}{\partial\boldsymbol{b}_i}$ by $-\frac{\sum_{k=1}^{K}f_{ik}^M\{T_i|M_i(T_i),\boldsymbol{w}_{ik};\boldsymbol{\theta}_{tk}\}}{1-\sum_{k=1}^{K}F_{ik}^M\{T_i|M_i(T_i),\boldsymbol{w}_{ik};\boldsymbol{\theta}_{tk}\}}$, which is always finite. Let $\boldsymbol{b}_i^{can}$ be a value for $\boldsymbol{b}_i$ proposed using the density $q(\boldsymbol{b}_i^{can}|\mathcal{D};\boldsymbol{\theta})\sim N\left\{\boldsymbol{b}_i^\star,(\boldsymbol{D}^{-1}+\omega\boldsymbol{Z}_{i,(m_i)}^\top\boldsymbol{Z}_{i,(m_i)})^{-1}\right\}$; note that the proposal distribution does not depend on the current value of $\boldsymbol{b}_i$, though it does depend on the current values of the remaining parameters, $\boldsymbol{\theta}_t$ and $\boldsymbol{\theta}_t$. The Metropolis-Hastings acceptance probability is therefore equal to

$$p=\min\left\{1,\frac{f(\boldsymbol{b}_i^{can}|\mathcal{D};\boldsymbol{\theta})}{f(\boldsymbol{b}_i|\mathcal{D};\boldsymbol{\theta})}\times\frac{q(\boldsymbol{b}_i|\mathcal{D};\boldsymbol{\theta})}{q(\boldsymbol{b}_i^{can}|\mathcal{D};\boldsymbol{\theta})}\right\}.$$

Thus, if the all-cause CIF is not bounded by 1 at $\boldsymbol{b}_i^{can}$, the acceptance probability is equal to zero as $f(\boldsymbol{b}_i^{can}|\mathcal{D};\boldsymbol{\theta})$ equals zero in this case.

5.2.3.2 Conditional posterior distribution of the survival parameters

The conditional posterior distribution of the survival model parameters, $\boldsymbol{\theta}_t$, is proportional to

$$f(\boldsymbol{\theta}_t|\mathcal{D},\boldsymbol{b};\boldsymbol{\beta}) \propto \prod_{i=1}^{N}\left(\prod_{k=1}^{K} f_{ik}^{M}\{T_i|M_i(T_i),\boldsymbol{w}_{ik};\boldsymbol{\theta}_{tk}\}^{\delta_{ik}}\left[1-\sum_{k=1}^{K}F_{ik}^{M}\{T_i|M_i(T_i),\boldsymbol{w}_{ik};\boldsymbol{\theta}_{tk}\}\right]^{1-\delta_i} \times I\left[\sum_{k=1}^{K}F_{ik}^{M}\{T_i|M_i(T_i),\boldsymbol{w}_{ik};\boldsymbol{\theta}_{tk}\}<1\right]\right) f(\boldsymbol{\theta}_t), \quad (5.8)$$

where $f(\boldsymbol{\theta}_t)$ is the prior distribution of $\boldsymbol{\theta}_t$. We updated the survival parameters $\boldsymbol{\theta}_{tk}$, $k=1,\ldots,K$, separately, performing BFGS updates (Thisted, 1988) maximising each distribution $f(\boldsymbol{\theta}_{tk}|\mathcal{D},\boldsymbol{b};\boldsymbol{\beta},\boldsymbol{\theta}_{sj},j\neq k)$, starting from the current value of the chain for $\boldsymbol{\theta}_{tk}$; note also that $f(\boldsymbol{\theta}_{tk}|\mathcal{D},\boldsymbol{b};\boldsymbol{\beta},\boldsymbol{\theta}_{sj},j\neq k)$ is proportional to (5.8). Thus, this method is based on the method proposed by Gamerman (1997). The BFGS algorithm can be briefly described as follows: Given an approximate information matrix $\boldsymbol{A}_k^{(j)}$, $j=0,1,\ldots$, we computed a new point

$$\boldsymbol{\theta}_{tk}^{(j+1)} = \boldsymbol{\theta}_{tk}^{(j)} + \{\boldsymbol{A}_k^{(j)}\}^{-1}\mathcal{U}\{\boldsymbol{\theta}_{tk}^{(j)}\},$$

where $\mathcal{U}(\boldsymbol{\theta}_{tk}) = \frac{\partial \log f(\boldsymbol{\theta}_{tk}|\mathcal{D},\boldsymbol{b};\boldsymbol{\beta},\boldsymbol{\theta}_{sj},j\neq k)}{\partial \boldsymbol{\theta}_{tk}}$ and $\boldsymbol{\theta}_{tk}^{(0)}$ is equal to the current value of the chain for $\boldsymbol{\theta}_{tk}$. The approximate information matrix $\boldsymbol{A}_k^{(j)}$ is also updated using the formula

$$\boldsymbol{A}_k^{(j+1)} = \boldsymbol{A}_k^{(j)} - \frac{\boldsymbol{y}\boldsymbol{y}^\top}{\boldsymbol{y}^\top\boldsymbol{\Sigma}} - \frac{\boldsymbol{A}_k^{(j)}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top\boldsymbol{A}_k^{(j)}}{\boldsymbol{\Sigma}^\top\boldsymbol{A}_k^{(j)}\boldsymbol{\Sigma}},$$

where $\boldsymbol{y} = \mathcal{U}\{\boldsymbol{\theta}_{tk}^{(j+1)}\} - \mathcal{U}\{\boldsymbol{\theta}_{tk}^{(j)}\}$ and $\boldsymbol{\Sigma} = \boldsymbol{\theta}_{tk}^{(j+1)} - \boldsymbol{\theta}_{tk}^{(j)}$. To find reasonably good approximations of the information matrices, before starting the MCMC algorithm, we maximized (5.8) using the estimates of the fixed and random effects from the LMM, resulting in an estimate, $\boldsymbol{\theta}_t^{LMM}$, and calculated the corresponding information matrix $\boldsymbol{A}$ at $\boldsymbol{\theta}_t^{LMM}$. Then we use the sub-matrix $\boldsymbol{A}_k$ associated with the parameter $\boldsymbol{\theta}_{tk}$ as the initial approximate information matrix $\boldsymbol{A}_k^{(0)}$. The main advantage of this procedure is that the computation of the hessian matrix of (5.8) is avoided, which reduces the computational burden. After performing the BFGS steps, we obtain a value $\boldsymbol{\theta}_{tk}^{\star}$. We

then propose a value $\boldsymbol{\theta}_{tk}^{can}$ from $q(\boldsymbol{\theta}_{tk}^{can}|\mathcal{D}, \boldsymbol{b}; \boldsymbol{\beta}, \boldsymbol{\theta}_{tk}, \boldsymbol{\theta}_{sj}, j \neq k) \sim t_{10}\left\{\boldsymbol{\theta}_{tk}^{\star}, (\boldsymbol{A}_k^{(0)})^{-1}\right\}$, where t_{10} denotes the multivariate Student-t distribution with location parameter and scale matrix equal to $\boldsymbol{\theta}_{tk}^{\star}$ and $(\boldsymbol{A}_k^{(0)})^{-1}$, respectively, using 10 degrees of freedom. The acceptance probability is equal to

$$p = \min\left\{1, \frac{f(\boldsymbol{\theta}_{tk}^{can}|\mathcal{D}, \boldsymbol{b}; \boldsymbol{\beta}, \boldsymbol{\theta}_{sj}, j \neq k)}{f(\boldsymbol{\theta}_{tk}|\mathcal{D}, \boldsymbol{b}; \boldsymbol{\beta}, \boldsymbol{\theta}_{sj}, j \neq k)} \times \frac{q(\boldsymbol{\theta}_{tk}|\mathcal{D}, \boldsymbol{b}; \boldsymbol{\beta}, \boldsymbol{\theta}_{tk}^{can}, \boldsymbol{\theta}_{sj}, j \neq k)}{q(\boldsymbol{\theta}_{tk}^{can}|\mathcal{D}, \boldsymbol{b}; \boldsymbol{\beta}, \boldsymbol{\theta}_{tk}, \boldsymbol{\theta}_{sj}, j \neq k)}\right\}.$$

We emphasize again that, if the proposed value for $\boldsymbol{\theta}_{tk}^{can}$ yields an all-cause CIF greater than 1, the acceptance probability is equal to 0. We also need to point out that, to calculate the proposal ratio, we need to perform the inverse BFGS step, starting from the proposed value, $\boldsymbol{\theta}_{tk}^{can}$, and recalculating the value of $\boldsymbol{\theta}_{tk}^{\star}$ in the numerator of the posterior ratio.

In our MCMC algorithm, we used fixed initial values $\boldsymbol{\beta}^{(0)}$, $\boldsymbol{D}^{(0)}$, and $\omega^{(0)}$ for $\boldsymbol{\beta}$, $\boldsymbol{D}$, and ω, respectively. The initial values for the random effects were simulated from $N\{\boldsymbol{b}_i^{LMM}, (\{\boldsymbol{D}^{(0)}\}^{-1} + \omega^{(0)}\boldsymbol{Z}_{i,(m_i)}^{\top}\boldsymbol{Z}_{i,(m_i)})^{-1}\}$, where $\boldsymbol{b}_i^{LMM}$ stands for the predictions of the random effects based on the LMM. Then we obtained initial values for $\boldsymbol{\theta}_t$ by maximising (5.8) given the initial values of $\{\boldsymbol{b}_i\}_{i=1}^N$ and $\boldsymbol{\beta}$. Under failure cause misclassification, after taking into account the true failure causes from doubly sampled patients, we obtained initial values for $\boldsymbol{\theta}_{misc}$ based on the full conditional posterior distribution of $\boldsymbol{\theta}_{misc}$.

5.2.3.3 Full conditional posterior distributions for the SREM-CIF-1 model

The posterior distribution of the random effects under the SREM-CIF-1 model is proportional to

$$\begin{aligned} f(\boldsymbol{b}_i|\mathcal{D}; \boldsymbol{\theta}) \propto \exp\Bigg[& -\frac{1}{2}\boldsymbol{b}_i^{\top}(\boldsymbol{D}^{-1} + \omega\boldsymbol{Z}_{i,(m_i)}^{\top}\boldsymbol{Z}_{i,(m_i)})\boldsymbol{b}_i + \omega\boldsymbol{b}_i^{\top}\boldsymbol{Z}_{i,(m_i)}^{\top}(\boldsymbol{y}_{i,(m_i)} - \boldsymbol{X}_{i,(m_i)}\boldsymbol{\beta}) \\ & + \sum_{k=1}^{K}\delta_{ik}\left\{\alpha_k\boldsymbol{b}_i^{\top}\boldsymbol{z}_i(T_i) - S_{ik}(\boldsymbol{b}_i)\right\} + (1-\delta_i)\log\left\{1 - K + \sum_{k=1}^{K}e^{-S_{ik}(\boldsymbol{b}_i)}\right\}\Bigg] \\ & \times I\left[\sum_{k=1}^{K}F_{ik}^{M}\{T_i|M_i(T_i), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_{tk}\} < 1\right] \end{aligned} \quad (5.9)$$

where $S_{ik}(\boldsymbol{b}_i) = \int_0^{T_i} e^{\boldsymbol{B}_k^\top(s)\boldsymbol{\psi}_k + \boldsymbol{\gamma}_k^\top \boldsymbol{w}_{ik} + \alpha_k m_i(s)} ds$. Note that (5.10) is equal to zero when the model-based CIF is not bounded by one due to the inclusion of the indicator function $I\left[\sum_{k=1}^K F_{ik}^M\{T_i|M_i(T_i), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_{tk}\} < 1\right]$. Ignoring the indicator function, the corresponding score function is equal to

$$\begin{aligned}\mathcal{U}(\boldsymbol{b}_i) &= -(\boldsymbol{D}^{-1} + \omega \boldsymbol{Z}_{i,(m_i)}^\top \boldsymbol{Z}_{i,(m_i)})\boldsymbol{b}_i + \omega \boldsymbol{Z}_{i,(m_i)}^\top (\boldsymbol{y}_{i,(m_i)} - \boldsymbol{X}_{i,(m_i)}\boldsymbol{\beta}) \\ &+ \sum_{k=1}^K \delta_{ik} \left\{ \alpha_k \boldsymbol{z}_i(T_i) - \frac{\partial S_{ik}(\boldsymbol{b}_i)}{\partial \boldsymbol{b}_i} \right\} - (1-\delta_i) \frac{\sum_{k=1}^K e^{-S_{ik}(\boldsymbol{b}_i)} \frac{\partial S_{ik}(\boldsymbol{b}_i)}{\partial \boldsymbol{b}_i}}{1 - K + \sum_{k=1}^K e^{-S_{ik}(\boldsymbol{b}_i)}},\end{aligned}$$

where $\frac{\partial S_{ik}(\boldsymbol{b}_i)}{\partial \boldsymbol{b}_i} = \int_0^{T_i} \alpha_k \boldsymbol{z}_i(s) e^{\boldsymbol{B}_k^\top(s)\boldsymbol{\psi}_k + \boldsymbol{\gamma}_k^\top \boldsymbol{w}_{ik} + \alpha_k m_i(s)} ds$. The information matrix is equal to

$$\mathcal{I}(\boldsymbol{b}_i) = (\boldsymbol{D}^{-1} + \omega \boldsymbol{Z}_{i,(m_i)}^\top \boldsymbol{Z}_{i,(m_i)}) + \sum_{k=1}^K \delta_{ik} \frac{\partial^2 S_{ik}(\boldsymbol{b}_i)}{\partial \boldsymbol{b}_i \boldsymbol{b}_i^\top} + (1-\delta_i) \frac{\partial}{\partial \boldsymbol{b}_i^\top} \left\{ \frac{\sum_{k=1}^K e^{-S_{ik}(\boldsymbol{b}_i)} \frac{\partial S_{ik}(\boldsymbol{b}_i)}{\partial \boldsymbol{b}_i}}{1 - K + \sum_{k=1}^K e^{-S_{ik}(\boldsymbol{b}_i)}} \right\},$$

where

$$\frac{\partial}{\partial \boldsymbol{b}_i^\top} \left\{ \frac{\sum_{k=1}^K e^{-S_{ik}(\boldsymbol{b}_i)} \frac{\partial S_{ik}(\boldsymbol{b}_i)}{\partial \boldsymbol{b}_i}}{1 - K + \sum_{k=1}^K e^{-S_{ik}(\boldsymbol{b}_i)}} \right\} = \boldsymbol{v}\boldsymbol{v}^\top +$$

$$\left\{ \sum_{k=1}^K e^{-S_{ik}(\boldsymbol{b}_i)} \frac{\partial^2 S_{ik}(\boldsymbol{b}_i)}{\partial \boldsymbol{b}_i \boldsymbol{b}_i^\top} - \sum_{k=1}^K e^{-S_{ik}(\boldsymbol{b}_i)} \frac{\partial S_{ik}(\boldsymbol{b}_i)}{\partial \boldsymbol{b}_i} \frac{\partial S_{ik}(\boldsymbol{b}_i)}{\partial \boldsymbol{b}_i^\top} \right\} \Big/ \left\{ 1 - K + \sum_{k=1}^K e^{-S_{ik}(\boldsymbol{b}_i)} \right\},$$

where $\boldsymbol{v} = \sum_{k=1}^K e^{-S_{ik}(\boldsymbol{b}_i)} \frac{\partial S_{ik}(\boldsymbol{b}_i)}{\partial \boldsymbol{b}_i} \Big/ \left\{ 1 - K + \sum_{k=1}^K e^{-S_{ik}(\boldsymbol{b}_i)} \right\}$ and

$$\frac{\partial^2 S_{ik}(\boldsymbol{b}_i)}{\partial \boldsymbol{b}_i \boldsymbol{b}_i^\top} = \int_0^{T_i} \alpha_k^2 \boldsymbol{z}_i(s) \boldsymbol{z}_i^\top(s) e^{\boldsymbol{B}_k^\top(s)\boldsymbol{\psi}_k + \boldsymbol{\gamma}_k^\top \boldsymbol{w}_{ik} + \alpha_k m_i(s)} ds.$$

The posterior distribution of the survival model parameters is proportional to

$$\begin{aligned} f(\boldsymbol{\theta}_t|\mathcal{D}, \boldsymbol{b}; \boldsymbol{\beta}) \quad &\propto \quad f(\boldsymbol{\theta}_t) \prod_{i=1}^N I\left[\sum_{k=1}^K F_{ik}^M\{T_i|M_i(T_i), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_{tk}\} < 1 \right] \\ &\times \quad \exp\left[\sum_{k=1}^K \delta_{ik} \left\{ \boldsymbol{B}_k^\top(T_i)\boldsymbol{\psi}_k + \boldsymbol{\gamma}_k^\top \boldsymbol{w}_{ik} + \alpha_k m_i(T_i) - S_{ik}(\boldsymbol{\theta}_{tk}) \right\} \right. \\ &+ \quad \left. (1-\delta_i) \log\left\{ 1 - K + \sum_{k=1}^K e^{-S_{ik}(\boldsymbol{\theta}_{tk})} \right\} \right], \end{aligned}$$

where $S_{ik}(\boldsymbol{\theta}_{tk}) = \int_0^{T_i} e^{\boldsymbol{B}_k^\top(s)\boldsymbol{\psi}_k + \boldsymbol{\gamma}_k^\top \boldsymbol{w}_{ik} + \alpha_k m_i(s)} ds$. The first-order partial derivatives are equal to

$$\frac{\partial \log f(\boldsymbol{\theta}_t|\mathcal{D}, \boldsymbol{b}; \boldsymbol{\beta})}{\partial \boldsymbol{\theta}_{tk}} = \frac{\partial \log f(\boldsymbol{\theta}_t)}{\partial \boldsymbol{\theta}_{tk}} + \sum_{i=1}^{N} \left[\delta_{ik} \left\{ \begin{pmatrix} \boldsymbol{B}_k(T_i) \\ \boldsymbol{w}_{ik} \\ m_i(T_i) \end{pmatrix} - \partial S_{ik}(\boldsymbol{\theta}_{tk}) / \partial \boldsymbol{\theta}_{tk} \right\} - (1 - \delta_i) \frac{e^{-S_{ik}(\boldsymbol{\theta}_{tk})} \partial S_{ik}(\boldsymbol{\theta}_{tk}) / \partial \boldsymbol{\theta}_{tk}}{1 - K + \sum_{k=1}^{K} e^{-S_{ik}(\boldsymbol{\theta}_{tk})}} \right],$$

where $\partial S_{ik}(\boldsymbol{\theta}_{tk}) / \partial \boldsymbol{\theta}_{tk} = \int_0^{T_i} \begin{pmatrix} \boldsymbol{B}_k(s) \\ \boldsymbol{w}_{ik} \\ m_i(s) \end{pmatrix} e^{\boldsymbol{B}_k^\top(s)\boldsymbol{\psi}_k + \boldsymbol{\gamma}_k^\top \boldsymbol{w}_{ik} + \alpha_k m_i(s)} ds.$

5.2.3.4 Full conditional posterior distributions for the SREM-CIF-2 model

The posterior distribution of the random effects under the SREM-CIF-2 model is proportional to

$$f(\boldsymbol{b}_i|\mathcal{D}; \boldsymbol{\theta}) \propto \exp\Bigg(-\frac{1}{2} \boldsymbol{b}_i^\top (\boldsymbol{D}^{-1} + \omega \boldsymbol{Z}_{i,(m_i)}^\top \boldsymbol{Z}_{i,(m_i)}) \boldsymbol{b}_i + \omega \boldsymbol{b}_i^\top \boldsymbol{Z}_{i,(m_i)}^\top (\boldsymbol{y}_{i,(m_i)} - \boldsymbol{X}_{i,(m_i)} \boldsymbol{\beta}) - \sum_{k=1}^{K} \delta_{ik} \frac{1 + c_k}{c_k} \log\{1 + c_k S_{ik}(\boldsymbol{b}_i)\} + \sum_{k=1}^{K} \delta_{ik} \alpha_k \boldsymbol{b}_i^\top \boldsymbol{z}_i(T_i) + (1 - \delta_i) \log\left[1 - K + \sum_{k=1}^{K} \{1 + c_k S_{ik}(\boldsymbol{b}_i)\}^{-1/c_k} \right] \Bigg) I\left[\sum_{k=1}^{K} F_{ik}^M \{T_i|M_i(T_i), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_{tk}\} < 1 \right] \quad (5.10)$$

The corresponding score function of (5.10) is equal to

$$\mathcal{U}(\boldsymbol{b}_i) = -(\boldsymbol{D}^{-1} + \omega \boldsymbol{Z}_{i,(m_i)}^\top \boldsymbol{Z}_{i,(m_i)}) \boldsymbol{b}_i + \omega \boldsymbol{Z}_{i,(m_i)}^\top (\boldsymbol{y}_{i,(m_i)} - \boldsymbol{X}_{i,(m_i)} \boldsymbol{\beta}) + \sum_{k=1}^{K} \left\{ -\delta_{ik} \frac{(1 + c_k) \frac{\partial S_{ik}(\boldsymbol{b}_i)}{\partial \boldsymbol{b}_i}}{1 + c_k S_{ik}(\boldsymbol{b}_i)} + \delta_{ik} \alpha_k \boldsymbol{z}_i(T_i) \right\} - (1 - \delta_i) \frac{\sum_{k=1}^{K} \{1 + c_k S_{ik}(\boldsymbol{b}_i)\}^{-(1+c_k)/c_k} \frac{\partial S_{ik}(\boldsymbol{b}_i)}{\partial \boldsymbol{b}_i}}{1 - K + \sum_{k=1}^{K} \{1 + c_k S_{ik}(\boldsymbol{b}_i)\}^{-1/c_k}}.$$

The information matrix is equal to

$$\mathcal{I}(\boldsymbol{b}_i) = \boldsymbol{D}^{-1} + \omega \boldsymbol{Z}_{i,(m_i)}^{\top} \boldsymbol{Z}_{i,(m_i)} + \sum_{k=1}^{K} \delta_{ik}(1+c_k) \frac{\frac{\partial^2 S_{ik}(\boldsymbol{b}_i)}{\partial \boldsymbol{b}_i \partial \boldsymbol{b}_i^{\top}} \left\{1 + c_k S_{ik}(\boldsymbol{b}_i)\right\} - c_k \frac{\partial S_{ik}(\boldsymbol{b}_i)}{\partial \boldsymbol{b}_i} \frac{\partial S_{ik}(\boldsymbol{b}_i)}{\partial \boldsymbol{b}_i^{\top}}}{\left\{1 + c_k S_{ik}(\boldsymbol{b}_i)\right\}^2}$$
$$(1-\delta_i) \frac{\partial}{\partial \boldsymbol{b}_i^{\top}} \left[\frac{\sum_{k=1}^{K} \left\{1 + c_k S_{ik}(\boldsymbol{b}_i)\right\}^{-(1+c_k)/c_k} \frac{\partial S_{ik}(\boldsymbol{b}_i)}{\partial \boldsymbol{b}_i}}{1 - K + \sum_{k=1}^{K} \left\{1 + c_k S_{ik}(\boldsymbol{b}_i)\right\}^{-1/c_k}} \right],$$

where

$$\frac{\partial}{\partial \boldsymbol{b}_i^{\top}} \left[\frac{\sum_{k=1}^{K} \left\{1 + c_k S_{ik}(\boldsymbol{b}_i)\right\}^{-(1+c_k)/c_k} \frac{\partial S_{ik}(\boldsymbol{b}_i)}{\partial \boldsymbol{b}_i}}{1 - K + \sum_{k=1}^{K} \left\{1 + c_k S_{ik}(\boldsymbol{b}_i)\right\}^{-1/c_k}} \right]$$
$$= \frac{\partial}{\partial \boldsymbol{b}_i^{\top}} \left[\sum_{k=1}^{K} \left\{1 + c_k S_{ik}(\boldsymbol{b}_i)\right\}^{-(1+c_k)/c_k} \frac{\partial S_{ik}(\boldsymbol{b}_i)}{\partial \boldsymbol{b}_i} \right] \Big/ \left[1 - K + \sum_{k=1}^{K} \left\{1 + c_k S_{ik}(\boldsymbol{b}_i)\right\}^{-1/c_k} \right] + \boldsymbol{v}\boldsymbol{v}^{\top}$$

where $\boldsymbol{v} = \sum_{k=1}^{K} \left\{1 + c_k S_{ik}(\boldsymbol{b}_i)\right\}^{-(1+c_k)/c_k} \frac{\partial S_{ik}(\boldsymbol{b}_i)}{\partial \boldsymbol{b}_i} \Big/ \left[1 - K + \sum_{k=1}^{K} \left\{1 + c_k S_{ik}(\boldsymbol{b}_i)\right\}^{-1/c_k}\right]$.
In addition,

$$\begin{aligned} &\frac{\partial}{\partial \boldsymbol{b}_i^{\top}} \sum_{k=1}^{K} \left\{1 + c_k S_{ik}(\boldsymbol{b}_i)\right\}^{-(1+c_k)/c_k} \frac{\partial S_{ik}(\boldsymbol{b}_i)}{\partial \boldsymbol{b}_i} = \\ - &\sum_{k=1}^{K} (1+c_k) \left\{1 + c_k S_{ik}(\boldsymbol{b}_i)\right\}^{-\frac{(1+c_k)}{c_k} - 1} \frac{\partial S_{ik}(\boldsymbol{b}_i)}{\partial \boldsymbol{b}_i} \frac{\partial S_{ik}(\boldsymbol{b}_i)}{\partial \boldsymbol{b}_i^{\top}} \\ + &\sum_{k=1}^{K} \left\{1 + c_k S_{ik}(\boldsymbol{b}_i)\right\}^{-(1+c_k)/c_k} \frac{\partial^2 S_{ik}(\boldsymbol{b}_i)}{\partial \boldsymbol{b}_i \partial \boldsymbol{b}_i^{\top}}. \end{aligned}$$

The conditional posterior distribution of the survival model parameters is proportional to

$$f(\boldsymbol{\theta}_t | \mathcal{D}, \boldsymbol{b}; \boldsymbol{\beta}) \propto f(\boldsymbol{\theta}_t) \prod_{i=1}^{N} \exp \left(\sum_{k=1}^{K} \left[-\delta_{ik} \frac{1+c_k}{c_k} \log \left\{1 + c_k S_{ik}(\boldsymbol{\theta}_{tk})\right\} \right.\right.$$
$$\left.\left. + \delta_{ik} \left\{ \boldsymbol{B}_k^{\top}(T_i) \boldsymbol{\psi}_k + \boldsymbol{\gamma}_k^{\top} \boldsymbol{w}_{ik} + \alpha_k m_i(T_i) \right\} \right] + (1-\delta_i) \log \left[1 - K + \sum_{k=1}^{K} \left\{1 + c_k S_{ik}(\boldsymbol{\theta}_{tk})\right\}^{-1/c_k} \right] \right)$$
$$\times I \left[\sum_{k=1}^{K} F_{ik}^{M} \left\{ T_i | M_i(T_i), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_{tk} \right\} < 1 \right],$$

with the score function being

$$\frac{\partial \log f(\boldsymbol{\theta}_t|\mathcal{D},\boldsymbol{b};\boldsymbol{\beta})}{\partial \boldsymbol{\theta}_{tk}} = \frac{\partial \log f(\boldsymbol{\theta}_t)}{\partial \boldsymbol{\theta}_{tk}} - \sum_{i=1}^{N} \delta_{ik} \frac{(1+c_k)\frac{\partial S_{ik}(\boldsymbol{\theta}_{tk})}{\partial \boldsymbol{\theta}_{tk}}}{1+c_k S_{ik}(\boldsymbol{\theta}_{tk})} + \sum_{i=1}^{N} \delta_{ik} \begin{pmatrix} \boldsymbol{B}_k(T_i) \\ \boldsymbol{w}_{ik} \\ m_i(T_i) \end{pmatrix} - \sum_{i=1}^{N} (1-\delta_i) \frac{\{1+c_k S_{ik}(\boldsymbol{\theta}_{tk})\}^{-(1+c_k)/c_k} \frac{\partial S_{ik}(\boldsymbol{\theta}_{tk})}{\partial \boldsymbol{\theta}_{tk}}}{1-K+\sum_{k=1}^{K}\{1+c_k S_{ik}(\boldsymbol{\theta}_{tk})\}^{-1/c_k}}.$$

5.3 Inference under potentially misclassified causes of failure

When the true failure cause, K_i, is not observed for all individuals, we assume that a cause of failure, $\tilde{K}_i$, is always reported although potentially misclassified (i.e. $K_i \neq \tilde{K}_i$). Let $\pi_{jk}(\mathcal{D}_{misc,i}) = \Pr(\tilde{K}_i = j|K_i = k, \mathcal{D}_{misc,i}; \boldsymbol{\theta}_{misc})$ be the probability of observing failure cause j given the true failure cause being k and some additional observed data $\mathcal{D}_{misc,i}$, with $\boldsymbol{\theta}_{misc}$ being the unknown parameter vector associated with $\pi_{jk}(\mathcal{D}_{misc,i})$. Note that $\mathcal{D}_{misc,i}$ may overlap with the observed covariates in the marker and survival models, but it may include additional auxiliary information that is not included in the models of interest. Thus, $\pi_{jk}(\mathcal{D}_{misc,i})$ allows for a differential misclassification mechanism (Daniel Paulino et al., 2003). In any case, we do not allow dependence of $\pi_{jk}(\mathcal{D}_{misc,i})$ on the random effects, $\boldsymbol{b}_i$. Note also that $\pi_{kk}(\mathcal{D}_{misc,i})$ is the probability of correctly classifying cause k, whereas $\sum_{j=1}^{K} \pi_{jk}(\mathcal{D}_{misc,i}) = 1$, for any $k \in \{1, 2, \ldots, K\}$. Moreover, we assume that non-informative right censoring is always correctly classified, i.e. $K_i = 0 \Leftrightarrow \tilde{K}_i = 0$ with probability 1.

We assume that the true failure cause is known in a small random sample of individuals, leading to a double sampling design (e.g. Bakoyannis et al., 2019). Let R_i be an indicator function of being doubly sampled (i.e. observing the true failure cause K_i on top of the potentially misclassified one, $\tilde{K}_i$). In this context, the full data are equal to $\mathcal{D}_{full} = \Big\{(\boldsymbol{Y}_{i,(m_i)}, \boldsymbol{X}_{i,(m_i)}, \boldsymbol{Z}_{i,(m_i)}, T_i, K_i, \tilde{K}_i, \boldsymbol{w}_i, \mathcal{D}_{misc,i}, \mathcal{D}_{R,i}, R_i), i = 1, 2, \ldots, N\Big\}$, where $\mathcal{D}_{R,i}$ denotes some potentially additional observed information related to the

probability of $R_i = 1$, whereas the ones actually observed are equal to

$$\mathcal{D}_{obs} = \begin{cases} (\boldsymbol{Y}_{i,(m_i)}, \boldsymbol{X}_{i,(m_i)}, \boldsymbol{Z}_{i,(m_i)}, T_i, K_i, \tilde{K}_i, \boldsymbol{w}_i, \mathcal{D}_{misc,i}, \mathcal{D}_{R,i}, R_i,), \ R_i = 1, \ i = 1, \ldots, N, \\ (\boldsymbol{Y}_{i,(m_i)}, \boldsymbol{X}_{i,(m_i)}, \boldsymbol{Z}_{i,(m_i)}, T_i, \tilde{K}_i, \boldsymbol{w}_i, \mathcal{D}_{misc,i}, \mathcal{D}_{R,i}, R_i), \ R_i = 0, \ i = 1, \ldots, N. \end{cases} . \tag{5.11}$$

Thus, using double sampling, the misclassification problem turns into a missing data problem, in the sense that the "true" failure causes for those who have failed but are not included in the double sample are missing. Under standard taxonomy of missing data mechanisms (Rubin, 1976), we make the MAR assumption, i.e. we assume that the probability of being included in the double sample depends on the observed data, but not on the missing true failure cause and the random effects. More formally, we assume that $\Pr(R_i = 1|\boldsymbol{Y}_{i,(m_i)}, \boldsymbol{X}_{i,(m_i)}, \boldsymbol{Z}_{i,(m_i)}, T_i, \tilde{K}_i, \boldsymbol{w}_i, \mathcal{D}_{misc,i}, \mathcal{D}_{R,i}; \boldsymbol{\theta}_R)$, i.e. the probability of being in the double sample contains all the observed information up to time T_i, with $\boldsymbol{\theta}_R$ being the parameter vector of the model. The MAR assumption, along with assuming prior independence between $(\boldsymbol{\theta}, \boldsymbol{\theta}_{misc})$ and $\boldsymbol{\theta}_R$, means that the model for $R = 1$ can be ignored for posterior inferences on $(\boldsymbol{\theta}, \boldsymbol{\theta}_{misc})$. In addition, it is also implied that the "true" failure cause can be validly predicted based on the observed data under the assumed model, i.e. $\Pr\{K_i = k|\tilde{K}_i = j, T_i^\star = t, M_i(t), \boldsymbol{w}_i, \mathcal{D}_{misc,i}; \boldsymbol{\theta}, \boldsymbol{\theta}_{misc}\} = \Pr\{K_i = k|\tilde{K}_i = j, T_i^\star = t, M_i(t), \boldsymbol{w}_i, \mathcal{D}_{misc,i}, \mathcal{D}_{R,i}, R_i; \boldsymbol{\theta}, \boldsymbol{\theta}_{misc}, \boldsymbol{\theta}_R\}$ for $R_i = 0, 1$. Then we show that the conditional posterior distribution of the "true" failure cause conditionally on the observed data is equal to

$$\Pr\{K_i = k|\tilde{K}_i = j, T_i^\star = t, M_i(t), \boldsymbol{w}_i, \mathcal{D}_{misc,i}; \boldsymbol{\theta}, \boldsymbol{\theta}_{misc}\} = \frac{f_{ik}\{t|M_i(t), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_{tk}\}\pi_{jk}(\mathcal{D}_{misc,i})}{\sum_{k=1}^{K} f_{ik}\{t|M_i(t), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_{tk}\}\pi_{jk}(\mathcal{D}_{misc,i})}. \tag{5.12}$$

To prove Equation (5.12), note that

$$\Pr\{K_i = k|\tilde{K}_i = j, T_i^\star = t, M_i(t), \boldsymbol{w}_i, \mathcal{D}_{misc,i}; \boldsymbol{\theta}, \boldsymbol{\theta}_{misc}\} = \frac{\Pr\{K_i = k, \tilde{K}_i = j|T_i^\star = t, M_i(t), \boldsymbol{w}_i, \mathcal{D}_{misc,i}; \boldsymbol{\theta}, \boldsymbol{\theta}_{misc}\}}{\Pr\{\tilde{K}_i = j|T_i^\star = t, M_i(t), \boldsymbol{w}_i, \mathcal{D}_{misc,i}; \boldsymbol{\theta}, \boldsymbol{\theta}_{misc}\}}$$

$$= \frac{\Pr\{K_i = k|T_i^\star = t, M_i(t), \boldsymbol{w}_i, \mathcal{D}_{misc,i}; \boldsymbol{\theta}, \boldsymbol{\theta}_{misc}\}\Pr\{\tilde{K}_i = j|K_i = k, T_i^\star = t, M_i(t), \boldsymbol{w}_i, \mathcal{D}_{misc,i}; \boldsymbol{\theta}, \boldsymbol{\theta}_{misc}\}}{\sum_{k=1}^{K}\Pr\{K_i = k|T_i^\star = t, M_i(t), \boldsymbol{w}_i, \mathcal{D}_{misc,i}; \boldsymbol{\theta}, \boldsymbol{\theta}_{misc}\}\Pr\{\tilde{K}_i = j|K_i = k, T_i^\star = t, M_i(t), \boldsymbol{w}_i, \mathcal{D}_{misc,i}; \boldsymbol{\theta}, \boldsymbol{\theta}_{misc}\}}$$

$$= \frac{\Pr\{K_i = k|T_i^\star = t, M_i(t), \boldsymbol{w}_i; \boldsymbol{\theta}_t\}\pi_{jk}(\mathcal{D}_{misc,i})}{\sum_{k=1}^{K}\Pr\{K_i = k|T_i^\star = t, M_i(t), \boldsymbol{w}_i; \boldsymbol{\theta}_t\}\pi_{jk}(\mathcal{D}_{misc,i})},$$

with the last result following from the model assumptions stating that (a) the failure cause probabilities do not depend on $\mathcal{D}_{misc,i}$ and $\boldsymbol{\theta}_{misc}$ and (b) the misclassification

probabilities $\pi_{jk}(\mathcal{D}_{misc,i})$ are independent of the random effects and the parameters of interest, $\boldsymbol{\theta}$, thus independent of $M_i(t)$ and $\boldsymbol{w}_i$. It is well known that the failure cause probabilities conditionally on the survival time $T_i^\star = t$ (Beyersmann et al., 2011) are equal to

$$\Pr\{K_i = k|T_i^\star = t, M_i(t), \boldsymbol{w}_i; \boldsymbol{\theta}_t\} = \frac{\alpha_{ik}\{t|M_i(t), \boldsymbol{w}_i; \boldsymbol{\theta}_t\}}{\sum_{k=1}^K \alpha_{ik}\{t|M_i(t), \boldsymbol{w}_i; \boldsymbol{\theta}_t\}},$$

where $\alpha_{ik}\{t|M_i(t), \boldsymbol{w}_i; \boldsymbol{\theta}_t\}$ denotes the kth cause-specific hazard function for individual i. Based on the CIF definition, we know that the CIF for cause k is equal to $F_{ik}\{t|M_i(t), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_{tk}\} = \int_0^t \alpha_{ik}\{u|M_i(u), \boldsymbol{w}_i; \boldsymbol{\theta}_t\} S_i\{u|M_i(u), \boldsymbol{w}_i; \boldsymbol{\theta}_t\} du$, which further implies that $f_{ik}\{t|M_i(t), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_{tk}\} = \alpha_{ik}\{t|M_i(t), \boldsymbol{w}_i; \boldsymbol{\theta}_t\} S_i\{t|M_i(t), \boldsymbol{w}_i; \boldsymbol{\theta}_t\}$. Therefore,

$$\Pr\{K_i = k|T_i^\star = t, M_i(t), \boldsymbol{w}_i; \boldsymbol{\theta}_t\} = \frac{f_{ik}\{t|M_i(t), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_{tk}\}}{\sum_{k=1}^K f_{ik}\{t|M_i(t), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_{tk}\}}.$$

Then it immediately follows that

$$\Pr\{K_i = k|\tilde{K}_i = j, T_i^\star = t, M_i(t), \boldsymbol{w}_i, \mathcal{D}_{misc,i}; \boldsymbol{\theta}, \boldsymbol{\theta}_{misc}\} = \frac{f_{ik}\{t|M_i(t), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_{tk}\}\pi_{jk}(\mathcal{D}_{misc,i})}{\sum_{k=1}^K f_{ik}\{t|M_i(t), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_{tk}\}\pi_{jk}(\mathcal{D}_{misc,i})}.$$

For individuals that are not doubly sampled, we only observe $\tilde{K}_i$, which may be different that K_i, though. For such cases, the observed survival data likelihood for cause j is equal to $f\{T_i, \tilde{K}_i = j|M_i(T_i), \boldsymbol{w}_i, \mathcal{D}_{misc,i}; \boldsymbol{\theta}_t, \boldsymbol{\theta}_{misc}\} = \sum_{k=1}^K f_{ik}\{t|M_i(t), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_{tk}\}\pi_{jk}(\mathcal{D}_{misc,i})$, i.e. a function involving both the density functions for all causes and the misclassification probabilities (Bakoyannis and Yiannoutsos, 2015). Thus, the observed data likelihood has a complicated form, which is difficult to handle. Similarly to Stamey et al. (2008) and Daniel Paulino et al. (2003), to deal with this issue, one can use data augmentation (Tanner and Wong, 1987), augmenting the observed likelihood for individuals who have failed from any event but are not included in the double sampling by the unobserved true failure causes, K_i. Letting $\mathcal{I}_{mis}$ be the indices for individuals that have failed from any event but are not doubly sampled: $\mathcal{I}_{mis} \equiv \{i : \tilde{K}_i \neq 0 \ \& \ R_i = 0\}$, the augmented posterior distribution of all unknown quantities is proportional to

$$f(\boldsymbol{\theta}, \boldsymbol{b}, \boldsymbol{\theta}_{misc}, \{K_i : i \in \mathcal{I}_{mis}\}|\mathcal{D}_{obs}) \propto f(\boldsymbol{\theta}) f(\boldsymbol{\theta}_{misc}) \prod_{i=1}^N \Bigg[f(\boldsymbol{Y}_{i,(m_i)}|; \boldsymbol{\theta}_L) f(\boldsymbol{b}_i; \boldsymbol{\theta}_L)$$
$$\prod_{k=1}^K f_{ik}\{T_i|M_i(T_i), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_{tk}\}^{\delta_{ik}} S_i\{T_i|M_i(T_i), \boldsymbol{w}_i; \boldsymbol{\theta}_t\}^{1-\delta_i} \prod_{j=1}^K \prod_{k=1}^K \pi_{jk}(\mathcal{D}_{misc,i})^{\tilde{\delta}_{ij}\delta_{ik}} \Bigg],$$

where we have actually factorized the full survival likelihood $f\{T_i, K_i, \tilde{K}_i | M_i(T_i), \boldsymbol{w}_i, \mathcal{D}_{misc,i}; \boldsymbol{\theta}_{tk}, \boldsymbol{\theta}_{misc}\}$ as the product of $f\{T_i, K_i | M_i(T_i), \boldsymbol{w}_i; \boldsymbol{\theta}_{tk}\}$ and $\Pr(\tilde{K}_i | K_i, \mathcal{D}_{misc,i}; \boldsymbol{\theta}_{misc})$. Note also that $\Pr(\tilde{K}_i = 0 | K_i = 0, \mathcal{D}_{misc,i}; \boldsymbol{\theta}_{misc}) = 1$, for the right censored individuals. It needs to be emphasized that data augmentation results in a much simpler form of the joint posterior distribution since, conditionally on the true failure causes K_i, the posterior distributions for $(\boldsymbol{\theta}, \boldsymbol{b})$ and $\boldsymbol{\theta}_{misc}$ are independent. In fact, the conditional posterior of $(\boldsymbol{\theta}, \boldsymbol{b})$ has the same form as in the case of no misclassification. The conditional posterior of $\{K_i : i \in \mathcal{I}_{mis}\}$ has been presented in (5.12). The following algorithm outlines the modified MCMC procedure to account for misclassification:

- Choose adequate initial values $\boldsymbol{\theta}^{(0)}, \boldsymbol{b}^{(0)}, \{K_i^{(0)} : i \in \mathcal{I}_{mis}\}, \boldsymbol{\theta}_{misc}^{(0)}$, meeting the likelihood constraints for all individuals.

- For $l = 1, 2, \ldots, L$

- Update $(\boldsymbol{\theta}^{(l-1)}, \boldsymbol{b}^{(l-1)})$ to $(\boldsymbol{\theta}^{(l)}, \boldsymbol{b}^{(l)})$ according to the posterior distribution $f(\boldsymbol{\theta}, \boldsymbol{b} | \{K_i^{(l-1)} : i \in \mathcal{I}_{mis}\}, \{K_i : i \notin \mathcal{I}_{mis}\}, \{(\boldsymbol{Y}_{i,(m_i)}, \boldsymbol{X}_{i,(m_i)}, \boldsymbol{Z}_{i,(m_i)}, T_i, \boldsymbol{w}_i), i = 1, \ldots, N\})$, i.e. the posterior distribution of the parameters of main interest, with the missing failure causes being equal to their current values. The MCMC algorithm for fully observed causes of failure described in Section 5.2 is used. However, the initial information matrix, $\boldsymbol{A}_k^{(0)}$, obtained by maximising the posterior distribution $f(\boldsymbol{\theta}_t | \mathcal{D}, \boldsymbol{b}^{LMM}; \boldsymbol{\beta}^{LMM})$ using the observed failure causes, may not be a good approximation. To address this issue, we perform a quick iterative procedure before starting the MCMC algorithm: (a) we maximize $f(\boldsymbol{\theta}_t | \mathcal{D}, \boldsymbol{b}^{LMM}; \boldsymbol{\beta}^{LMM})$ using simulated missing failure causes $\{K_i : i \in \mathcal{I}_{mis}\}$ and (b) based on the current mode of $f(\boldsymbol{\theta}_t | \mathcal{D}, \boldsymbol{b}^{LMM}; \boldsymbol{\beta}^{LMM})$ we carry out an approximate Gibbs sampling procedure of 100 iterations for the conditional posterior distribution of $(\boldsymbol{\theta}_{misc}, \{K_i : i \in \mathcal{I}_{mis}\})$ by repeatedly simulating the missing failure causes and locating the mode of the conditional posterior distribution of $\boldsymbol{\theta}_{misc}$ given the simulated failure causes $\{K_i : i \in \mathcal{I}_{mis}\}$. This approach is repeated 5 times and $\boldsymbol{A}_k^{(0)}$ corresponds to the information matrix for $\boldsymbol{\theta}_{tk}$ at the last iteration.

- Update $\boldsymbol{\theta}_{misc}^{(l-1)}$ to $\boldsymbol{\theta}_{misc}^{(l)}$ according to $f(\boldsymbol{\theta}_{misc}|\{K_i^{(l-1)} : i \in \mathcal{I}_{mis}\}, \{K_i : i \notin \mathcal{I}_{mis}\}, \{\mathcal{D}_{misc,i}, i = 1, \ldots, N\})$.
- Sample $\{K_i^{(l)} : i \in \mathcal{I}_{mis}\}$ from $f\{K_i|\tilde{K}_i, T_i, M_i^{(l)}(T_i), \boldsymbol{w}_i, \mathcal{D}_{misc,i}; \boldsymbol{\theta}^{(l)}, \boldsymbol{\theta}_{misc}^{(l)}\}$ directly using Equation (5.12).

To model the misclassification probabilities, we use Multinomial logistic regression updating the values of $\boldsymbol{\theta}_{misc}$ using the approach proposed by Gamerman (1997). The conditional posterior distribution of the misclassification parameters is proportional to

$$f(\boldsymbol{\theta}_{misc}|\{(K_i, \mathcal{D}_{misc,i})\}_{i=1}^N) \propto f(\boldsymbol{\theta}_{misc}) \prod_{i=1}^{N} \prod_{j=1}^{K} \prod_{k=1}^{K} \pi_{jk}(\mathcal{D}_{misc,i})^{\tilde{\delta}_{ij}\delta_{ik}}.$$

For brevity, we present in detail the case of $K = 2$ competing risks, which corresponds to a logistic regression model:

$$\Pr(\tilde{K}_i = 1|K_i, \mathcal{D}_{misc,i}; \boldsymbol{\theta}_{misc}) = \frac{e^{\boldsymbol{X}_{misc,i}^\top \boldsymbol{\theta}_{misc}}}{1 + e^{\boldsymbol{X}_{misc,i}^\top \boldsymbol{\theta}_{misc}}} = p_i, \tag{5.13}$$

where $\boldsymbol{X}_{misc,i}$ denotes the relevant design matrix associating K_i and $\mathcal{D}_{misc,i}$ with the probability of observing $\tilde{K}_i = 1$. Thus, $\Pr(\tilde{K}_i = 2|K_i, \mathcal{D}_{misc,i}; \boldsymbol{\theta}_{misc}) = 1 - p_i = \{1 + e^{\boldsymbol{X}_{misc,i}^\top \boldsymbol{\theta}_{misc}}\}^{-1}$. Ignoring, at this stage, the prior distribution, the conditional posterior distribution of $\boldsymbol{\theta}_{misc}$ is proportional to $\prod_{i=1}^N p_i^{\tilde{\delta}_{i1}}(1 - p_i)^{\tilde{\delta}_{i2}}$. The posterior distribution on the log scale is, thus, equal to

$$\log f(\boldsymbol{\theta}_{misc}|\{(K_i, \mathcal{D}_{misc,i})\}_{i=1}^N) = \sum_{i=1}^{N} \left\{ \tilde{\delta}_{i1} \boldsymbol{X}_{misc,i}^\top \boldsymbol{\theta}_{misc} - \delta_i \log(1 + e^{\boldsymbol{X}_{misc,i}^\top \boldsymbol{\theta}_{misc}}) \right\}.$$

The score vector and the information matrix of $\boldsymbol{\theta}_{misc}$ are equal to

$$\mathcal{U}(\boldsymbol{\theta}_{misc}) = \sum_{i=1}^{N} \left\{ \tilde{\delta}_{i1} \boldsymbol{X}_{misc,i} - \delta_i p_i \boldsymbol{X}_{misc,i} \right\}$$

and $\mathcal{I}(\boldsymbol{\theta}_{misc}) = \sum_{i=1}^N \delta_i \boldsymbol{X}_{misc,i} p_i (1 - p_i) \boldsymbol{X}_{misc,i}^\top$, respectively. To update the values of $\boldsymbol{\theta}_{misc}$, we used the method proposed by Gamerman (1997) by carrying out one Newton-Raphson step starting at the current value and proposing a value from a multivariate Normal distribution with mean obtained from the Newton-Raphson step and covariance matrix the inverse of the information matrix at the current value.

Alternatively, a standard special case occurs when one does not intend to include covariate information in $\pi_{jk}(\mathcal{D}_{misc,i})$, i.e. $\mathcal{D}_{misc,i}$ is empty. Then, a natural choice for the prior distributions of π_{jk}'s would be the independent Dirichlet$(a_{1k}, \ldots, a_{Kk})$ distributions, i.e. $f(\pi_{1k}, \ldots, \pi_{Kk}) \propto \prod_{j=1}^{K} \pi_{jk}^{a_{jk}-1}$, for $k = 1, \ldots, K$. It can be shown that the posterior distributions of $\pi_{1k}, \ldots, \pi_{Kk}$ are then independent Dirichlet$(a_{1k} + \sum_{i=1}^{N} \tilde{\delta}_{i1}\delta_{ik}, a_{2k} + \sum_{i=1}^{N} \tilde{\delta}_{i2}\delta_{ik}, \ldots, a_{Kk} + \sum_{i=1}^{N} \tilde{\delta}_{iK}\delta_{ik})$, $k = 1, 2, \ldots, K$. For $K = 2$ competing risks, there are two distinct misclassification parameters. Assuming Beta(a_1, b_1) and Beta(a_2, b_2) prior distributions for π_{11} and π_{22}, respectively, results in the corresponding conditional posterior distributions being Beta$(a_1 + \sum_{i=1}^{N} \tilde{\delta}_{i1}\delta_{i1}, b_1 + \sum_{i=1}^{N} \tilde{\delta}_{i2}\delta_{i1})$ and Beta$(a_2 + \sum_{i=1}^{N} \tilde{\delta}_{i2}\delta_{i2}, b_2 + \sum_{i=1}^{N} \tilde{\delta}_{i1}\delta_{i2})$.

5.4 Posterior inferences for population-averaged CIFs and marker states

To describe the cohort evolution over time, states defined by marker data and clinical outcomes are often used. A pragmatic approach to do so is to discretize the marker values into non-overlapping intervals $\{[s_0, s_1), \ldots, [s_{J-1}, s_J)\}$ and define mutually-exclusive states based on clinical events and (discretized) marker data. If the focus of the analysis lies in describing the "true" biological process, as often is the case in the joint modeling literature, states may be defined in terms of the "true" marker values, i.e., for any $t > 0$, $\{m_i(t) \in S_h, T_i^\star > t\}$, $h = 1, \ldots, J$ and $\{T_i^\star \leq t, K_i = k\}$, $k = 1, \ldots, K$, where $S_h = [s_{h-1}, s_h)$. Progression of the whole cohort can be easily monitored by a series of estimated multistate probabilities $\Pr\{m_i(t) \in S_h, T_i^\star > t | \boldsymbol{w}_i; \boldsymbol{\theta}\}$, $h = 1, \ldots, J$ and $\Pr(T_i^\star \leq t, K_i = k | \boldsymbol{w}_{ik}; \boldsymbol{\theta})$, $k = 1, \ldots, K$, through a multistate probability plot. The first quantity, referred to as latent marker state probability, expresses the probability of being event free and having true marker values in S_h. The second expression, i.e. $\Pr(T_i^\star \leq t, K_i = k | \boldsymbol{w}_{ik}; \boldsymbol{\theta})$, is the population-averaged CIF for a particular cause. To get better insight into the dynamics of the processes, one may be also interested in transitions between states. In real-life applications (e.g. Stover et al., 2019), for simplicity,

transitions are often defined from baseline states. Letting $p_g(0) = \Pr\{m_i(0) \in S_g; \boldsymbol{\theta}_L\}$, it can be easily shown that

$$\Pr\{T_i^\star \leq t, K_i = k | m_i(0) \in S_g, \boldsymbol{w}_{ik}; \boldsymbol{\theta}\} = \int\limits_{m_i(0) \in S_g} F_{ik}\{t | M_i(t), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_{tk}\} \frac{f(\boldsymbol{b}_i; \boldsymbol{\theta}_L)}{p_g(0)} d\boldsymbol{b}_i \tag{5.14}$$

$$\Pr\{m_i(t) \in S_h, T_i^\star > t | m_i(0) \in S_g, \boldsymbol{w}_i; \boldsymbol{\theta}\} = \int\limits_{m_i(0) \in S_g, m_i(t) \in S_h} S_i\{t | M_i(t), \boldsymbol{w}_i; \boldsymbol{\theta}_t\} \frac{f(\boldsymbol{b}_i; \boldsymbol{\theta}_L)}{p_g(0)} d\boldsymbol{b}_i \tag{5.15}$$

Inference on (5.14) and (5.15) involves two distinct problems (i) approximation of the integral over the random effects and (ii) accounting for the variability in $\boldsymbol{\theta}$.

5.4.1 Estimation procedure

We initially describe the estimation of (5.14) and (5.15) for any given $\boldsymbol{\theta}$. Specifically, (5.14) can be approximated by drawing samples $\{\boldsymbol{b}_{ig}^{(j)}\}_{j=1}^{N\text{mc}}$ for $\boldsymbol{b}_i$ from the $N(\boldsymbol{0}, \boldsymbol{D})$ distribution under the linear constraint $m_i(0) \in S_g$, which can be carried out, among many other options, through Hamiltonian Monte Carlo (Pakman, 2015). Specifically,

$$\Pr\{T_i^\star \leq t, K_i = k | m_i(0) \in S_g, \boldsymbol{w}_{ik}; \boldsymbol{\theta}\} \approx N_{\text{mc}}^{-1} \sum_{j=1}^{N\text{mc}} F_{ik}\{t | M_{ig}^{(j)}(t), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_{tk}\} \tag{5.16}$$

where $m_{ig}^{(j)}(t) = \boldsymbol{x}_i^\top(t)\boldsymbol{\beta} + \boldsymbol{z}_i^\top(t)\boldsymbol{b}_{ig}^{(j)}$ and $M_{ig}^{(j)}(t) = \{m_{ig}^{(j)}(s) : 0 \leq s \leq t\}$. Using similar ideas, it can be seen that, after multiplying and dividing (5.15) by $\Pr\{m_i(t) \in S_h, m_i(0) \in S_g; \boldsymbol{\theta}\}$, (5.15) can be approximated using samples $\{\boldsymbol{b}_{igh}^{(j)}\}_{j=1}^{N\text{mc}}$ from the $N(\boldsymbol{0}, \boldsymbol{D})$ distribution under the linear constraints $m_i(0) \in S_g$ and $m_i(t) \in S_h$, i.e. $\Pr\{m_i(t) \in S_h, T_i^\star > t | m_i(0) \in S_g, \boldsymbol{w}_i; \boldsymbol{\theta}\}$ can be approximated by

$$\frac{\Pr\{m_i(t) \in S_h, m_i(0) \in S_g; \boldsymbol{\theta}\}}{\Pr\{m_i(0) \in S_g; \boldsymbol{\theta}\} N_{\text{mc}}} \sum_{j=1}^{N\text{mc}} S_i\{t | M_{igh}^{(j)}(t), \boldsymbol{w}_i; \boldsymbol{\theta}_t\}, \tag{5.17}$$

where $m_{igh}^{(j)}(t) = \boldsymbol{x}_i^\top(t)\boldsymbol{\beta} + \boldsymbol{z}_i^\top(t)\boldsymbol{b}_{igh}^{(j)}$ and $M_{igh}^{(j)}(t) = \{m_{igh}^{(j)}(s) : 0 \leq s \leq t\}$. Since $m_i(t) \sim N\{\boldsymbol{x}_i^\top(t)\boldsymbol{\beta}, \boldsymbol{z}_i^\top(t)\boldsymbol{D}\boldsymbol{z}_i(t)\}$, $\Pr\{m_i(t) \in S_h, m_i(0) \in S_g; \boldsymbol{\theta}\}$ and $\Pr\{m_i(0) \in S_g; \boldsymbol{\theta}\}$ can be

easily computed using the cumulative distribution function of the (bivariate) Normal distribution. Due to (5.2), it should be noted that if $\sum_{k=1}^{K} F_{ik}^{M}\{t|M_{ig}^{(j)}(t), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_{tk}\} > 1$, $F_{ik}\{t|M_{ig}^{(j)}(t), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_{tk}\} = F_{ik}\{t'|M_{ig}^{(j)}(t'), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_{tk}\}$, where $t' = \tau_i(\boldsymbol{\beta}, \boldsymbol{\theta}_t, \boldsymbol{b}_{ig}^{(j)})$, thus calculation of the upper bound is required only for the random draws that do not fulfil the boundedness constraint.

A posterior sample for (5.14) and (5.15) can be obtained by (a) drawing $\boldsymbol{\theta}^{(l)} \sim f(\boldsymbol{\theta}|\mathcal{D}_{\text{obs}})$, $l = 1, 2, \ldots, L$ and (b) approximating $\Pr\{m_i(t) \in S_h, T_i^{\star} > t|m_i(0) \in S_g, \boldsymbol{w}_i; \boldsymbol{\theta}^{(l)}\}$ and $\Pr\{m_i(t) \in S_h, T_i^{\star} > t|m_i(0) \in S_g, \boldsymbol{w}_i; \boldsymbol{\theta}^{(l)}\}$, for each $l = 1, 2, \ldots, L$, using (5.16) and (5.17). Thus, posterior means and posterior credible intervals can easily be estimated. Also, once posterior samples for (5.14) and (5.15) are available, it is easy to get posterior samples for population-averaged CIFs and latent marker state probabilities through the following relationships $\Pr(T_i^{\star} \leq t, K_i = k|\boldsymbol{w}_{ik}; \boldsymbol{\theta}) = \sum_{g=1}^{J} \Pr\{T_i^{\star} \leq t, K_i = k|m_i(0) \in S_g, \boldsymbol{w}_{ik}; \boldsymbol{\theta}\} \Pr\{m_i(0) \in S_g; \boldsymbol{\theta}\}$ and $\Pr\{m_i(t) \in S_h, T_i^{\star} > t|\boldsymbol{w}_i; \boldsymbol{\theta}\} = \sum_{g=1}^{J} \Pr\{m_i(t) \in S_h, T_i^{\star} > t|m_i(0) \in S_g, \boldsymbol{w}_i; \boldsymbol{\theta}\} \Pr\{m_i(0) \in S_g; \boldsymbol{\theta}\}$, respectively.

In theory, $\sum_{k=1}^{K} \Pr\{T_i^{\star} \leq t, K_i = k|m_i(0) \in S_g, \boldsymbol{w}_{ik}; \boldsymbol{\theta}\}$ is equal to $1 - \sum_{h=1}^{J} \Pr\{m_i(t) \in S_h, T_i^{\star} > t|m_i(0) \in S_g, \boldsymbol{w}_i; \boldsymbol{\theta}\}$. However, due to Monte Carlo approximation error, results using (5.14) and (5.15) might differ slightly. To get consistent results, we used $1 - \sum_{k=2}^{K} \Pr\{T_i^{\star} \leq t, K_i = k|m_i(0) \in S_g, \boldsymbol{w}_{ik}; \boldsymbol{\theta}^{(l)}\} - \sum_{k=2}^{J} \Pr\{m_i(t) \in S_k, T_i^{\star} > t|m_i(0) \in S_g, \boldsymbol{w}_i; \boldsymbol{\theta}^{(l)}\}$ as the posterior sample for $\Pr\{T_i^{\star} \leq t, K_i = 1|m_i(0) \in S_g, \boldsymbol{w}_{i1}; \boldsymbol{\theta}\}$, $g = 1, \ldots, J$.

5.4.2 CIF estimates conditional on observed marker states

In a clinical application, the population-averaged CIF conditional on the observed marker state would be valuable for prediction purposes, as it is the only available information at baseline. Thus, for prediction purposes, one could be also interested in $\Pr\{T_i^{\star} \leq t, K_i = k|y_i(0) \in S_g, \boldsymbol{w}_{ik}; \boldsymbol{\theta}\}$, $g = 1, \ldots, J$. By standard calculations, it can be shown that $\Pr\{T_i^{\star} \leq t, K_i = k|y_i(0) \in S_g, \boldsymbol{w}_{ik}; \boldsymbol{\theta}\}$ is equal to

$$\Pr\{T_i^\star \le t, K_i = k | y_i(0) \in S_g, \boldsymbol{w}_{ik}; \boldsymbol{\theta}\} = \frac{\Pr\{T_i^\star \le t, K_i = k, y_i(0) \in S_g | \boldsymbol{w}_{ik}; \boldsymbol{\theta}\}}{\Pr\{y_i(0) \in S_g; \boldsymbol{\theta}\}}$$

$$= \int\limits_{y_i(0) \in S_g} \int_0^t \frac{f\{u, k, y_i(0) | \boldsymbol{w}_{ik}; \boldsymbol{\theta}\}}{\Pr\{y_i(0) \in S_g; \boldsymbol{\theta}\}} \, du \, dy_i(0)$$

$$= \int\limits_{y_i(0) \in S_g} \int_0^t \int \frac{f\{u, k, y_i(0) | \boldsymbol{b}_i, \boldsymbol{w}_{ik}; \boldsymbol{\theta}\} f(\boldsymbol{b}_i; \boldsymbol{\theta})}{\Pr\{y_i(0) \in S_g; \boldsymbol{\theta}\}} \, d\boldsymbol{b}_i \, du \, dy_i(0).$$

Due to the model assumption of conditional independence between the marker and survival process given the random effects, $\Pr\{T_i^\star \le t, K_i = k | y_i(0) \in S_g, \boldsymbol{w}_{ik}; \boldsymbol{\theta}\}$ is equal to

$$\int\limits_{y_i(0) \in S_g} \int \left[\int_0^t f(u, k | \boldsymbol{b}_i, \boldsymbol{w}_{ik}; \boldsymbol{\theta}) \, du \right] \frac{f\{y_i(0) | \boldsymbol{b}_i; \boldsymbol{\theta}\} f(\boldsymbol{b}_i; \boldsymbol{\theta})}{\Pr\{y_i(0) \in S_g; \boldsymbol{\theta}\}} \, d\boldsymbol{b}_i \, dy_i(0)$$

$$= \int\limits_{y_i(0) \in S_g} \int F_{ik}\{t | M_i(t), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_{tk}\} \frac{f\{y_i(0), \boldsymbol{b}_i; \boldsymbol{\theta}\}}{\Pr\{y_i(0) \in S_g; \boldsymbol{\theta}\}} \, d\boldsymbol{b}_i \, dy_i(0) \tag{5.18}$$

Thus, (5.18) can be estimated by drawing samples $\{y_{ig}^{(j)}(0), \boldsymbol{b}_{ig}^{(j)}\}_{j=1}^{N_{\text{mc}}}$ for $\{y_i(0), \boldsymbol{b}_i\}$ from the

$$N\left\{ \begin{pmatrix} \boldsymbol{x}_i^\top(0)\boldsymbol{\beta} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \sigma^2 + \boldsymbol{z}_i^\top(0)\boldsymbol{D}\boldsymbol{z}_i(0) & \boldsymbol{z}_i^\top(0)\boldsymbol{D} \\ \boldsymbol{D}\boldsymbol{z}_i(0) & \boldsymbol{D} \end{pmatrix} \right\},$$

distribution, constrained such that $y_i(0) \in S_g$, i.e. $\Pr\{T_i^\star \le t, K_i = k | y_i(0) \in S_g, \boldsymbol{w}_{ik}; \boldsymbol{\theta}\}$ can be approximated by $N_{\text{mc}}^{-1} \sum_{j=1}^{N_{\text{mc}}} F_{ik}\{t | M_{ig}^{(j)}(t), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_{tk}\}$, where $m_{ig}^{(j)}(t) = \boldsymbol{x}_i^\top(t)\boldsymbol{\beta} + \boldsymbol{z}_i^\top(t)\boldsymbol{b}_{ig}^{(j)}$ and $M_{ig}^{(j)}(t) = \{m_{ig}^{(j)}(s) : 0 \le s \le t\}$.

Moreover, one may be also interested in CIFs conditional on being in certain states at specific time points. In this case, it would be reasonable to condition on survival up to the last time point, i.e. $\Pr\{T_i^\star \le t, K_i = k | T_i^\star > s, y_i(0) \in S_g, y_i(s) \in S_h, \boldsymbol{w}_{ik}; \boldsymbol{\theta}\}$, for $0 \le s < t$ and $g, h \in \{1, 2, \ldots, J\}$. These probabilities can be estimated in a similar

way. Specifically,

$$\begin{aligned}
&\frac{\Pr\{s < T_i^\star \le t, K_i = k, y_i(0) \in S_g, y_i(s) \in S_h | \boldsymbol{w}_i; \boldsymbol{\theta}\}}{\Pr\{s < T_i^\star, y_i(0) \in S_g, y_i(s) \in S_h | \boldsymbol{w}_i; \boldsymbol{\theta}\}} \\
&= \frac{\int\limits_{y_i(0)\in S_g} \int\limits_{y_i(s)\in S_h} \int_s^t f\{u, k, y_i(0), y_i(s) | \boldsymbol{w}_i; \boldsymbol{\theta}\}\, du\, dy_i(s)\, dy_i(0)}{\int\limits_{y_i(0)\in S_g} \int\limits_{y_i(s)\in S_h} \int_s^\infty f\{u, y_i(0), y_i(s) | \boldsymbol{w}_i; \boldsymbol{\theta}\}\, du\, dy_i(s)\, dy_i(0)} \\
&= \frac{\int\limits_{y_i(0)\in S_g} \int\limits_{y_i(s)\in S_h} \int_s^t \int f\{u, k, y_i(0), y_i(s) | \boldsymbol{b}_i, \boldsymbol{w}_i; \boldsymbol{\theta}\} f(\boldsymbol{b}_i; \boldsymbol{\theta})\, d\boldsymbol{b}_i\, du\, dy_i(s)\, dy_i(0)}{\int\limits_{y_i(0)\in S_g} \int\limits_{y_i(s)\in S_h} \int_s^\infty \int f\{u, y_i(0), y_i(s) | \boldsymbol{b}_i, \boldsymbol{w}_i; \boldsymbol{\theta}\} f(\boldsymbol{b}_i; \boldsymbol{\theta})\, d\boldsymbol{b}_i\, du\, dy_i(s)\, dy_i(0)}
\end{aligned}$$

By the conditional independence of the proposed model, i.e. the marker and survival processes are independent given the random effects, it follows that $\Pr\{T_i^\star \le t, K_i = k | T_i^\star > s, y_i(0) \in S_g, y_i(s) \in S_h, \boldsymbol{w}_{ik}; \boldsymbol{\theta}\}$ is equal to

$$\begin{aligned}
&\frac{\int\limits_{y_i(0)\in S_g} \int\limits_{y_i(s)\in S_h} \int_s^t \int f(u, k | \boldsymbol{b}_i, \boldsymbol{w}_i; \boldsymbol{\theta}) f\{y_i(0), y_i(s) | \boldsymbol{b}_i; \boldsymbol{\theta}\} f(\boldsymbol{b}_i; \boldsymbol{\theta})\, d\boldsymbol{b}_i\, du\, dy_i(s)\, dy_i(0)}{\int\limits_{y_i(0)\in S_g} \int\limits_{y_i(s)\in S_h} \int_s^\infty \int f(u | \boldsymbol{b}_i, \boldsymbol{w}_i; \boldsymbol{\theta}) f\{y_i(0), y_i(s) | \boldsymbol{b}_i; \boldsymbol{\theta}\} f(\boldsymbol{b}_i; \boldsymbol{\theta})\, d\boldsymbol{b}_i\, du\, dy_i(s)\, dy_i(0)} \\
&= \frac{\int\limits_{y_i(0)\in S_g} \int\limits_{y_i(s)\in S_h} \int \left[\int_s^t f(u, k | \boldsymbol{b}_i, \boldsymbol{w}_i; \boldsymbol{\theta})\, du\right] f\{y_i(0), y_i(s) | \boldsymbol{b}_i; \boldsymbol{\theta}\} f(\boldsymbol{b}_i; \boldsymbol{\theta})\, d\boldsymbol{b}_i\, dy_i(s)\, dy_i(0)}{\int\limits_{y_i(0)\in S_g} \int\limits_{y_i(s)\in S_h} \int \left[\int_s^\infty f(u | \boldsymbol{b}_i, \boldsymbol{w}_i; \boldsymbol{\theta})\, du\right] f\{y_i(0), y_i(s) | \boldsymbol{b}_i; \boldsymbol{\theta}\} f(\boldsymbol{b}_i; \boldsymbol{\theta})\, d\boldsymbol{b}_i\, dy_i(s)\, dy_i(0)} \\
&= \frac{\int\limits_{y_i(0)\in S_g} \int\limits_{y_i(s)\in S_h} \int \left[F_{ik}\{t | M_i(t), \boldsymbol{w}_i; \boldsymbol{\theta}_{tk}\} - F_{ik}\{s | M_i(s), \boldsymbol{w}_i; \boldsymbol{\theta}_{tk}\}\right] \frac{f\{y_i(0), y_i(s), \boldsymbol{b}_i; \boldsymbol{\theta}\}}{\Pr\{y_i(0)\in S_g, y_i(s)\in S_h; \boldsymbol{\theta}\}}\, d\boldsymbol{b}_i\, dy_i(s)\, dy_i(0)}{\int\limits_{y_i(0)\in S_g} \int\limits_{y_i(s)\in S_h} \int S_i\{s | M_i(s), \boldsymbol{w}_i; \boldsymbol{\theta}_t\} \frac{f\{y_i(0), y_i(s), \boldsymbol{b}_i; \boldsymbol{\theta}\}}{\Pr\{y_i(0)\in S_g, y_i(s)\in S_h; \boldsymbol{\theta}\}}\, d\boldsymbol{b}_i\, dy_i(s)\, dy_i(0)}
\end{aligned}$$

Thus, in a very similar way, $\Pr\{T_i^\star \le t, K_i = k | T_i^\star > s, y_i(0) \in S_g, y_i(s) \in S_h, \boldsymbol{w}_i; \boldsymbol{\theta}\}$ can be estimated by drawing samples $\{y_{igh}^{(j)}(0), y_{igh}^{(j)}(s), \boldsymbol{b}_{igh}^{(j)}\}_{j=1}^{N\text{mc}}$ for $\{y_i(0), y_i(s), \boldsymbol{b}_i\}$ from the

$$N\left\{\begin{pmatrix} \boldsymbol{x}_i^\top(0)\boldsymbol{\beta} \\ \boldsymbol{x}_i^\top(s)\boldsymbol{\beta} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} \sigma^2 + \boldsymbol{z}_i^\top(0)\boldsymbol{D}\boldsymbol{z}_i(0) & \boldsymbol{z}_i^\top(0)\boldsymbol{D}\boldsymbol{z}_i(s) & \boldsymbol{z}_i(0)^\top\boldsymbol{D} \\ \boldsymbol{z}_i^\top(0)\boldsymbol{D}\boldsymbol{z}_i(s) & \sigma^2 + \boldsymbol{z}_i^\top(s)\boldsymbol{D}\boldsymbol{z}_i(s) & \boldsymbol{z}_i(s)^\top\boldsymbol{D} \\ \boldsymbol{D}\boldsymbol{z}_i(0) & \boldsymbol{D}\boldsymbol{z}_i(s) & \boldsymbol{D} \end{pmatrix}\right\},$$

distribution, constrained such that $y_i(0) \in S_g$ and $y_i(s) \in S_h$, i.e. $\Pr\{T_i^\star \le t, K_i =$

$k|T_i^\star > s, y_i(0) \in S_g, y_i(s) \in S_h, \boldsymbol{w}_i; \boldsymbol{\theta}\}$ can be approximated by

$$\frac{\sum_{j=1}^{N\text{mc}} \left[F_{ik}\{t|M_{igh}^{(j)}(t), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_{tk}\} - F_{ik}\{s|M_{igh}^{(j)}(s), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_{tk}\}\right]}{\sum_{j=1}^{N\text{mc}} S_i\{s|M_{igh}^{(j)}(s), \boldsymbol{w}_i; \boldsymbol{\theta}_t\}},$$

where $m_{igh}^{(j)}(t) = \boldsymbol{x}_i^\top(t)\boldsymbol{\beta} + \boldsymbol{z}_i^\top(t)\boldsymbol{b}_{igh}^{(j)}$ and $M_{igh}^{(j)}(t) = \{m_{igh}^{(j)}(s) : 0 \leq s \leq t\}$.

5.5 Comparison of models' fit using the DIC criterion

To compare the fit of models SREM-CIF-1 and SREM-CIF-2, we use the marginalised version (Quintero and Lesaffre, 2018) of the deviance information criterion (DIC) proposed by Spiegelhalter et al. (2002a). To compute the marginal DIC criterion, we need to calculate the observed data likelihood, which requires integration over the random effects. Assuming that there is no misclassification, the observed data likelihood is equal to

$$\begin{aligned} f(\mathcal{D}; \boldsymbol{\theta}) &= \prod_{i=1}^{N} f(\boldsymbol{Y}_{i,(m_i)}, T_i, K_i|\boldsymbol{w}_i; \boldsymbol{\theta}) \\ &= \prod_{i=1}^{N} f(\boldsymbol{Y}_{i,(m_i)}; \boldsymbol{\theta}) \int f\{T_i, K_i|M_i(T_i), \boldsymbol{w}_i; \boldsymbol{\theta}_t\} f(\boldsymbol{b}_i|\boldsymbol{Y}_{i,(m_i)}; \boldsymbol{\theta}) d\boldsymbol{b}_i, \end{aligned} \quad (5.19)$$

where we have actually factorized $f(\boldsymbol{Y}_{i,(m_i)}, T_i, K_i|\boldsymbol{w}_i; \boldsymbol{\theta})$ as $f(\boldsymbol{Y}_{i,(m_i)}; \boldsymbol{\theta}) f(T_i, K_i|\boldsymbol{Y}_{i,(m_i)}, \boldsymbol{w}_i; \boldsymbol{\theta})$, with $f(T_i, K_i|\boldsymbol{Y}_{i,(m_i)}, \boldsymbol{w}_i; \boldsymbol{\theta})$ being equal to $\int f\{T_i, K_i|M_i(T_i), \boldsymbol{w}_i; \boldsymbol{\theta}_t\} f(\boldsymbol{b}_i|\boldsymbol{Y}_{i,(m_i)}; \boldsymbol{\theta}) d\boldsymbol{b}_i$, due to the conditional independence assumption for $\boldsymbol{Y}_{i,(m_i)}|\boldsymbol{b}_i$ and $T_i^\star|\boldsymbol{b}_i$. Also, it is straightforward to show that $\boldsymbol{Y}_{i,(m_i)}; \boldsymbol{\theta} \sim N(\boldsymbol{X}_{i,(m_i)}\boldsymbol{\beta}, \omega^{-1}\boldsymbol{I}_{m_i} + \boldsymbol{Z}_{i,(m_i)}\boldsymbol{D}\boldsymbol{Z}_{i,(m_i)}^\top)$ and $\boldsymbol{b}_i|\boldsymbol{Y}_{i,(m_i)}; \boldsymbol{\theta} \sim N(\boldsymbol{\mu}_{\boldsymbol{b}_i}, \boldsymbol{\Sigma}_{\boldsymbol{b}_i})$, where $\boldsymbol{\Sigma}_{\boldsymbol{b}_i} = (\boldsymbol{D}^{-1} + \omega\boldsymbol{Z}_{i,(m_i)}^\top\boldsymbol{Z}_{i,(m_i)})^{-1}$ and $\boldsymbol{\mu}_{\boldsymbol{b}_i} = \boldsymbol{\Sigma}_{\boldsymbol{b}_i}\omega\boldsymbol{Z}_{i,(m_i)}^\top(\boldsymbol{Y}_{i,(m_i)} - \boldsymbol{X}_{i,(m_i)}\boldsymbol{\beta})$. Letting $Dev(\boldsymbol{\theta}) = -2\log f(\mathcal{D}; \boldsymbol{\theta})$ be the deviance of the model, the DIC criterion is defined as $\text{DIC} = Dev\{E(\boldsymbol{\theta}|\mathcal{D})\} + 2p_{Eff}$, where $Dev\{E(\boldsymbol{\theta}|\mathcal{D})\}$ is the deviance evaluated at the posterior mean of the parameters and p_{Eff} is the effective number of parameters. Based on Spiegelhalter et al. (2002a), p_{Eff} should be equal to $p_{Eff} = E\{Dev(\boldsymbol{\theta})|\mathcal{D}\} - Dev\{E(\boldsymbol{\theta}|\mathcal{D})\}$, the posterior mean of the deviance minus the deviance at the posterior mean of the parameters. $E\{Dev(\boldsymbol{\theta})|\mathcal{D}\}$ can be approximated by the sample mean of the deviances evaluated at each value of

the MCMC algorithm, while $Dev\{E(\boldsymbol{\theta}|\mathcal{D})\}$ by the deviance evaluated at the mean of the MCMC sample.

In theory, DIC would be straightforward to approximate based on the MCMC sample if the deviance was available in closed form, but the integral in (5.19) cannot be calculated analytically. Similarly to the pseudo-adaptive quadrature method proposed by Rizopoulos (2012b), to consistently estimate $f(T_i, K_i|\boldsymbol{Y}_{i,(m_i)}, \boldsymbol{w}_i; \boldsymbol{\theta})$, we used Monte Carlo integration, i.e. $f(T_i, K_i|\boldsymbol{Y}_{i,(m_i)}, \boldsymbol{w}_i; \boldsymbol{\theta})$ is approximated by $N_{mc}^{-1} f\{T_i, K_i|M_i^{(j)}(t), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_t\}$, where $m_i^{(j)}(t) = \boldsymbol{x}_i^\top(t)\boldsymbol{\beta} + \boldsymbol{z}_i^\top(t)\boldsymbol{b}_i^{(j)}$, where $\boldsymbol{b}_i^{(j)} \sim N(\boldsymbol{\mu}_{\boldsymbol{b}_i}, \boldsymbol{\Sigma}_{\boldsymbol{b}_i})$, $j = 1, \ldots, N_{mc}$. The above Monte Carlo estimator had great performance as $\boldsymbol{\mu}_{\boldsymbol{b}_i}$ was very close to the mode of the integrand in (5.19) in almost all cases, a remark consistent with the literature (Rizopoulos, 2012b). However, in very few cases, the model-based all-cause CIF may be greater than 1 when evaluated at $\boldsymbol{\mu}_{\boldsymbol{b}_i}$, leading to $f\{T_i, K_i|M_i(T_i), \boldsymbol{w}_i; \boldsymbol{\theta}_t\}$ being equal to zero when evaluated at $\boldsymbol{\mu}_{\boldsymbol{b}_i}$. In this case, the variance of the Monte Carlo estimator could be too large. To overcome this issue, we used importance sampling by exploiting the posterior sample of $\boldsymbol{b}_i$ available from MCMC. Specifically, we (a) simulate $\boldsymbol{b}_i^{(j)} \sim N(\bar{\boldsymbol{\mu}}_{\boldsymbol{b}_i}, \boldsymbol{\Sigma}_{\boldsymbol{b}_i})$, $j = 1, \ldots, N_{mc}$, where $\bar{\boldsymbol{\mu}}_{\boldsymbol{b}_i}$ is the marginal posterior mean of the random effect, $\boldsymbol{b}_i$, and (b) approximate $f(T_i, K_i|\boldsymbol{Y}_{i,(m_i)}, \boldsymbol{w}_i; \boldsymbol{\theta})$ by $N_{mc}^{-1} f\{T_i, K_i|M_i^{(j)}(t), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_t\} f(\boldsymbol{b}_i^{(j)}|\boldsymbol{Y}_{i,(m_i)}; \boldsymbol{\theta})/g(\boldsymbol{b}_i^{(j)})$, where $g(\boldsymbol{b}_i^{(j)})$ is the density of the $N(\bar{\boldsymbol{\mu}}_{\boldsymbol{b}_i}, \boldsymbol{\Sigma}_{\boldsymbol{b}_i})$ distribution at $\boldsymbol{b}_i^{(j)}$ and $M_i^{(j)}(t) = \{m_i^{(j)}(s) : 0 \leq s \leq t\}$, with $m_i^{(j)}(t) = \boldsymbol{x}_i^\top(t)\boldsymbol{\beta} + \boldsymbol{z}_i^\top(t)\boldsymbol{b}_i^{(j)}$, $j = 1, \ldots, N_{mc}$. Since the deviance has to be computed at each MCMC iteration, we cannot know beforehand if the importance sampling estimator would perform better than the simple Monte Carlo estimator. Thus, we employ the importance sampling estimator only if $\bar{\boldsymbol{\mu}}_{\boldsymbol{b}_i}$ leads to higher posterior density, $f(\boldsymbol{b}_i|\boldsymbol{Y}_{i,(m_i)}, T_i, K_i; \boldsymbol{\theta})$, than $\boldsymbol{\mu}_{\boldsymbol{b}_i}$ does. Note that the conditional posterior distribution $f(\boldsymbol{b}_i|\boldsymbol{Y}_{i,(m_i)}, T_i, K_i; \boldsymbol{\theta})$ is proportional to the integrand in (5.19).

To evaluate the DIC criterion under failure cause misclassification, some minor adjustments are required, since the true failure causes are not available for all individuals. We have actually used exactly the same procedure but the term $f\{T_i, K_i|M_i(T_i), \boldsymbol{w}_i; \boldsymbol{\theta}_t\}$ in (5.19) should be replaced by $f\{T_i, K_i|M_i(T_i), \boldsymbol{w}_i; \boldsymbol{\theta}_t\} \times \prod_{j=1}^{K} \prod_{k=1}^{K} \pi_{jk}(\mathcal{D}_{misc,i})^{\tilde{\delta}_{ij}\delta_{ik}}$

and $\sum_{j=1}^{K} \tilde{\delta}_{ij} \sum_{k=1}^{K} f_{ik}\{t|M_i(t), \boldsymbol{w}_{ik}; \boldsymbol{\theta}_{tk}\}\pi_{jk}(\mathcal{D}_{misc,i})$,, for doubly and non-doubly sampled patients, respectively. For the right-censored individuals, nothing changed as $\Pr(\tilde{K}_i = 0|K_i = 0, \mathcal{D}_{misc}, \boldsymbol{\theta}_{misc}) = 1$.

5.6 Simulation study

5.6.1 Simulation study under no misclassification of failure cause

A simulation study was carried out to evaluate the performance of the proposed methodology under certain conditions. Marker data were generated using a piece-wise linear LMM $y_i(t) = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1})\min(t, 1) + (\beta_2 + b_{i2})[\max\{\min(t, 5), 1\} - 1] + \beta_3\{\max(t, 5) - 5\} + \epsilon_i(t)$, with $(b_{i0}, b_{i1}, b_{i2}) \sim N(\mathbf{0}, \boldsymbol{D})$ and $\epsilon_i(t) \sim N(0, \omega^{-1})$. Thus, the population slopes are β_1, β_2, and β_3, when $t \in [0, 1)$, $t \in [1, 5)$, and $t > 5$, respectively. Note that this model mimics, at least roughly, the CD4 cell count evolution since ART initiation. Measurements were assumed to be collected biannually and the maximum study duration was assumed to be 10 years. The true values of $\boldsymbol{\beta}$, vech($\boldsymbol{D}$), and ω were assumed to be $(12.85, 6.03, 0.77, 0)^\top$, $(25.09, -5.08, -2.65, 10.18, 1.09, 0.85)^\top$ and $1/8.14$, respectively, where vech stands for the "vector-half" operator stacking the columns of the lower triangular part of a symmetric matrix. We assumed 2 competing risks (e.g. $K_i = 1, 2$ corresponding to death and disengagement from care, respectively), with the marker measurements after the first occurring event being ignored. Two scenarios regarding the competing-risk data were considered: survival data were simulated based on (a) the SREM-CIF-1 model and (b) the SREM-CIF-2 model ($c_1 = c_2 = 1$), according to the following equations:

$$F_{ik}^{M}\{t|M_i(t), w_i; \boldsymbol{\theta}_{tk}\} = 1 - \exp\left\{-\int_0^t u_{k1}(s)e^{\gamma_k w_i + \alpha_k m_i(s)}ds\right\}, \quad \textbf{SREM-CIF-1}$$

$$F_{ik}^{M}\{t|M_i(t), w_i; \boldsymbol{\theta}_{tk}\} = 1 - \left\{1 + \int_0^t u_{k2}(s)e^{\gamma_k w_i + \alpha_k m_i(s)}ds\right\}^{-1}, \quad \textbf{SREM-CIF-2}$$

where $k = 1, 2$ and w_i is a binary baseline covariate following the Bernoulli distribution with probability of success equal to 0.5. For each simulation scenario, we simulated 500 datasets, each including $N = 1500$ individuals. To generate the competing-risk

data, based on Beyersmann et al. (2011), we first simulated the overall survival time $T_i^\star$ within $(0, \tau_i)$ using the inverse CDF theorem and then determined the failure causes with corresponding probabilities equal to

$$\frac{\partial F_{ik}^M\{T_i^\star|M_i(T_i^\star), w_i; \boldsymbol{\theta}_{tk}\}/\partial T_i^\star \partial F_{i1}^M\{T_i^\star|M_i(T_i^\star), w_i; \boldsymbol{\theta}_{s1}\}}{\partial T_i^\star + \partial F_{i2}^M\{T_i^\star|M_i(T_i^\star), w_i; \boldsymbol{\theta}_{s2}\}/\partial T_i^\star},$$

$k = 1, 2$. An independent right censoring mechanism was also applied using $C_i \sim \min(U_i, 10)$, where $U_i \sim \text{Exp}(0.025)$ (i.e. the exponential distribution with rate=0.025). Regarding the true parameter values, we assumed that $\alpha_1 = -0.16$, $\alpha_2 = -0.02$, $\gamma_1 = 0.15$, and $\gamma_2 = -0.15$. For the baseline levels, we assumed that $u_{11}(t) = 0.62 \times 0.25\exp(-0.25t + 7 \times 0.16)/[1 - 0.62\{1 - \exp(-0.25t)\}]$, $u_{21}(t) = 0.70 \times 0.13\exp(-0.13t + 7 \times 0.02)/[1 - 0.70\{1 - \exp(-0.13t)\}]$, $u_{12}(t) = 0.67 \times 0.25\exp(-0.25t + 7 \times 0.16)/[1 - 0.67\{1 - \exp(-0.25t)\}]$, and $u_{22}(t) = 0.83 \times 0.13\exp(-0.13t + 7 \times 0.02)/[1 - 0.83\{1 - \exp(-0.13t)\}]$.

Under each of the two scenarios for the survival submodels, we fitted the proposed model using both the SREM-CIF-1 and SREM-CIF-2 parameterisations. This way, we can evaluate the sensitivity of the results to misspecification of the link function for the CIF submodels. The marker model was correctly specified in the fitted models though. The B-splines matrices $\boldsymbol{B}_k(t)$ approximating the baseline CIF levels had 2 and 3 knots for the first and the second failure type, respectively, placed at the observed quantiles of the respective event times. To derive inference on the model parameters, we applied the MCMC algorithm described in subsection 5.2.3 using 200 draws as a burn-in period and recorded additional 10500 iterations. To account for the autocorrelation in the MCMC sample we thinned the chain by keeping every third draw, thus producing posterior inferences based on 3500 draws. For the population parameters and the survival parameters, $\boldsymbol{\beta}$ and $\boldsymbol{\theta}_t$, respectively, we assumed independent normal distributions with zero means and variances equal to 100, for the covariance matrix of the random effects, $\boldsymbol{D}$, we assumed the Inverse-Wishart $IW(\boldsymbol{A}, df)$ distribution with $df = 3$ and $\boldsymbol{A} = 3 \times \text{diag}(25, 5, 5)$, and for the within-subject precision, ω, the Gamma$(0.01, 0.01)$ distribution. Apart from the model parameters, we examined the performance of the proposed approach in deriving inferences on (a) latent marker state probabilities over

time, (b) transition probabilities by baseline marker states, (c) population-averaged CIFs, and (d) population-averaged CIFs by baseline marker states, using the methodology described in Section 5.4. These estimates were produced at times 0, 2, 4, 6, 8, and 10 years. The numerical integration required was performed through $N_{mc} = 1000$ Monte Carlo draws using a posterior sample for $\boldsymbol{\theta}$ of $L = 350$ draws, obtained by keeping every tenth MCMC draw. Parameter estimates were based on the posterior medians, with the corresponding credible intervals estimated by the observed 2.5% and 97.5% quantile of the MCMC posterior sample. For each fitted model, we calculated the marginal DIC criterion using $N_{mc} = 200$, based on 500 draws from the posterior distribution of $\boldsymbol{\theta}$ obtained by keeping every seventh draw. The performance of the DIC criterion at correctly identifying the true model was assessed by recording the proportion of time the true model was chosen under both scenarios.

To assess model performance, we present the bias, the Monte carlo standard deviation, the average model-based standard error, and the empirical coverage probability of the respective credible intervals. Since parameters γ_k and α_k do not have the same interpretation under SREM-CIF-1 and SREM-CIF-2, we did not provide bias and coverage probability results for γ_k and α_k when the fitted model was misspecified. The results under the SREM-CIF-1 and SREM-CIF-2 scenarios are presented in Tables 5.1 and 5.2, respectively. The fixed-effect estimates were approximately unbiased for both models under the two scenarios (bias: from -0.011 to 0.010), while the coverage probabilities were close to the nominal level (from 92.6 to 96.4%). The estimates for γ_k and α_k were nearly unbiased along with approximately 95% coverage probabilities when the fitted model coincided with the true data generating mechanism. The DIC criterion had moderate ability to identify the correct model as it selected the true model 75.0% and 62.4% of the time under the SREM-CIF-1 and SREM-CIF-2 scenarios, respectively. Its discriminating ability substantially increased to 90.0% and 86.0%, respectively, when a simulation study including 8000 individuals and 50 replications was performed. Focusing on the population CIFs estimates, both models yielded estimates with negligible bias along with adequate empirical coverage probabilities, while

the Monte Carlo standard deviation was close to the average model-based standard error. Thus, misspecification of the link function of the survival submodels does not seem to affect the performance of population CIFs estimates.

Table 5.1: Simulation study results from fitted SREM-CIF-1 and SREM-CIF-2 models when the data have been generated by the SREM-CIF-1 model†.

Parameter	True	Median	Bias	ASD	MCSD	Coverage	Median	Bias	ASD	MCSD	Coverage
Longitudinal	Results from SREM-CIF-1						Results from SREM-CIF-2				
Intercept	12.850	12.851	0.001	0.145	0.156	92.600	12.851	0.001	0.145	0.156	93.000
Slope1 (β_1)	6.030	6.026	-0.004	0.126	0.134	92.600	6.019	-0.011	0.126	0.133	92.600
Slope2 (β_2)	0.770	0.768	-0.002	0.036	0.036	95.400	0.766	-0.004	0.036	0.036	96.200
Slope3 (β_3)	0.000	-0.001	-0.001	0.019	0.020	94.800	-0.002	-0.002	0.019	0.020	94.600
Cause1 (e.g. death)											
"True" marker value (α_1)	-0.160	-0.161	-0.001	0.014	0.014	95.800	-0.179		0.017	0.016	
Binary covariate (γ_1)	0.150	0.145	-0.005	0.132	0.137	93.400	0.157		0.148	0.153	
CIF1 $t = 2$, $w = 1$	8.175	8.093	-0.082	0.838	0.818	93.600	8.120	-0.055	0.854	0.827	95.200
CIF1 $t = 4$, $w = 1$	11.533	11.397	-0.135	1.039	1.016	95.000	11.428	-0.105	1.056	1.020	95.200
CIF1 $t = 6$, $w = 1$	13.495	13.365	-0.130	1.157	1.129	95.800	13.389	-0.106	1.171	1.126	95.600
CIF1 $t = 8$, $w = 1$	14.774	14.640	-0.134	1.230	1.187	95.400	14.661	-0.113	1.242	1.186	95.600
CIF1 $t = 10$, $w = 1$	15.604	15.455	-0.148	1.278	1.247	95.600	15.473	-0.131	1.288	1.247	95.600
CIF1 $t = 2$, $w = 0$	7.106	7.073	-0.033	0.774	0.789	93.600	7.092	-0.015	0.790	0.807	93.600
CIF1 $t = 4$, $w = 0$	10.062	9.995	-0.067	0.976	0.984	94.200	10.038	-0.024	0.992	1.000	94.800
CIF1 $t = 6$, $w = 0$	11.799	11.743	-0.056	1.094	1.097	95.000	11.799	0.001	1.108	1.109	95.600
CIF1 $t = 8$, $w = 0$	12.935	12.880	-0.056	1.167	1.166	95.000	12.948	0.013	1.179	1.172	95.200
CIF1 $t = 10$, $w = 0$	13.673	13.607	-0.067	1.216	1.212	94.200	13.684	0.010	1.226	1.219	95.400
Cause2 (e.g. disengagement)											
"True" marker value (α_2)	-0.020	-0.021	-0.001	0.009	0.009	94.600	-0.026		0.011	0.011	
Binary covariate (γ_2)	-0.150	-0.149	0.001	0.086	0.087	94.400	-0.180		0.105	0.106	
CIF2 $t = 2$, $w = 1$	11.525	11.494	-0.030	0.914	0.869	94.800	11.412	-0.113	0.950	0.909	95.200
CIF2 $t = 4$, $w = 1$	20.352	20.290	-0.062	1.294	1.290	95.000	20.249	-0.102	1.338	1.342	94.200
CIF2 $t = 6$, $w = 1$	27.328	27.258	-0.071	1.540	1.581	94.200	27.293	-0.035	1.565	1.619	93.200
CIF2 $t = 8$, $w = 1$	32.937	32.864	-0.073	1.711	1.747	94.000	32.991	0.054	1.709	1.762	93.800
CIF2 $t = 10$, $w = 1$	37.431	37.352	-0.080	1.839	1.899	92.600	37.566	0.135	1.811	1.880	92.200
CIF2 $t = 2$, $w = 0$	13.260	13.212	-0.048	1.003	1.009	93.400	13.356	0.096	1.047	1.059	94.400
CIF2 $t = 4$, $w = 0$	23.228	23.138	-0.090	1.379	1.406	93.400	23.297	0.069	1.424	1.452	94.200
CIF2 $t = 6$, $w = 0$	30.982	30.868	-0.114	1.610	1.620	95.600	30.978	-0.004	1.631	1.631	94.400
CIF2 $t = 8$, $w = 0$	37.130	37.014	-0.116	1.765	1.789	93.800	37.057	-0.073	1.756	1.765	94.400
CIF2 $t = 10$, $w = 0$	41.997	41.861	-0.136	1.876	1.902	93.800	41.834	-0.163	1.838	1.846	94.000

† Results from 500 replications with each dataset including 1500 individuals. The true marker evolution was based on linear splines with knots at 1 and 5 years since baseline and it was correctly specified in the fitted SREM-CIF-1 and SREM-CIF-2 models. "True" denotes the true parameter values; "Median" the mean of posterior medians over the 500 replications; "Bias" the mean bias for posterior median estimates; "ASD" the average posterior standard deviation, "MCSD" the empirical Monte carlo deviation of estimates and "Coverage" the empirical coverage probability of posterior credible intervals.

Table 5.2: Simulation study results from fitted SREM-CIF-1 and SREM-CIF-2 models when the data have been generated by the SREM-CIF-2 model.

Parameter	True	Median	Bias	ASD	MCSD	Coverage	Median	Bias	ASD	MCSD	Coverage
Longitudinal	Results from SREM-CIF-1						Results from SREM-CIF-2				
Intercept	12.850	12.845	-0.005	0.146	0.137	95.600	12.844	-0.006	0.146	0.138	95.800
Slope1 (β_1)	6.030	6.040	0.010	0.127	0.126	94.600	6.033	0.003	0.127	0.126	94.600
Slope2 (β_2)	0.770	0.772	0.002	0.036	0.034	95.600	0.770	-0.000	0.036	0.034	96.400
Slope3 (β_3)	0.000	0.001	0.001	0.019	0.020	92.600	0.001	0.001	0.019	0.020	92.600
Cause1 (e.g. death)											
"True" marker value (α_1)	-0.160	-0.143		0.014	0.014		-0.161	-0.001	0.017	0.016	96.000
Binary covariate (γ_1)	0.150	0.133		0.133	0.132		0.146	-0.004	0.149	0.148	94.200
CIF1 $t = 2$, $w = 1$	8.319	8.216	-0.103	0.853	0.841	95.000	8.270	-0.048	0.869	0.858	93.800
CIF1 $t = 4$, $w = 1$	11.572	11.464	-0.108	1.055	1.052	94.400	11.529	-0.043	1.069	1.064	93.400
CIF1 $t = 6$, $w = 1$	13.469	13.376	-0.093	1.172	1.171	95.600	13.439	-0.030	1.182	1.180	94.000
CIF1 $t = 8$, $w = 1$	14.709	14.608	-0.102	1.245	1.235	96.000	14.669	-0.040	1.251	1.244	95.200
CIF1 $t = 10$, $w = 1$	15.521	15.430	-0.091	1.296	1.288	95.400	15.489	-0.032	1.298	1.294	95.200
CIF1 $t = 2$, $w = 0$	7.288	7.258	-0.030	0.794	0.774	95.400	7.274	-0.014	0.808	0.786	95.400
CIF1 $t = 4$, $w = 0$	10.196	10.158	-0.039	0.995	0.977	94.600	10.198	0.002	1.007	0.990	94.600
CIF1 $t = 6$, $w = 0$	11.904	11.871	-0.033	1.112	1.094	93.800	11.920	0.017	1.120	1.102	94.200
CIF1 $t = 8$, $w = 0$	13.027	12.979	-0.049	1.185	1.180	94.400	13.041	0.013	1.190	1.181	94.400
CIF1 $t = 10$, $w = 0$	13.765	13.723	-0.042	1.235	1.265	93.600	13.791	0.026	1.238	1.264	93.800
Cause2 (e.g. disengagement)											
"True" marker value (α_2)	-0.020	-0.015		0.009	0.009		-0.019	0.001	0.011	0.011	95.400
Binary covariate (γ_2)	-0.150	-0.128		0.086	0.086		-0.157	-0.007	0.105	0.104	94.200
CIF2 $t = 2$, $w = 1$	12.861	12.924	0.063	0.984	0.981	94.800	12.826	-0.035	1.020	1.019	94.000
CIF2 $t = 4$, $w = 1$	21.826	21.839	0.014	1.352	1.321	94.000	21.766	-0.060	1.389	1.364	94.200
CIF2 $t = 6$, $w = 1$	28.518	28.472	-0.046	1.576	1.583	94.000	28.457	-0.061	1.593	1.600	94.200
CIF2 $t = 8$, $w = 1$	33.726	33.594	-0.132	1.728	1.781	93.000	33.647	-0.079	1.717	1.770	93.400
CIF2 $t = 10$, $w = 1$	37.837	37.586	-0.252	1.842	1.932	92.600	37.700	-0.137	1.805	1.894	92.400
CIF2 $t = 2$, $w = 0$	14.636	14.547	-0.089	1.060	1.105	93.200	14.685	0.049	1.106	1.143	93.000
CIF2 $t = 4$, $w = 0$	24.488	24.406	-0.082	1.422	1.438	93.200	24.545	0.057	1.463	1.461	94.200
CIF2 $t = 6$, $w = 0$	31.666	31.640	-0.026	1.633	1.640	94.000	31.738	0.072	1.650	1.646	94.200
CIF2 $t = 8$, $w = 0$	37.150	37.172	0.022	1.774	1.804	94.400	37.217	0.068	1.760	1.778	94.600
CIF2 $t = 10$, $w = 0$	41.417	41.438	0.022	1.877	1.908	94.400	41.427	0.011	1.834	1.854	93.800

† Results from 500 replications with each dataset including 1500 individuals. The true marker evolution was based on linear splines with knots at 1 and 5 years since baseline and it was correctly specified in the fitted SREM-CIF-1 and SREM-CIF-2 models. "True" denotes the true parameter values; "Median" the mean of posterior medians over the 500 replications; "Bias" the mean bias for posterior median estimates; "ASD" the average posterior standard deviation, "MCSD" the empirical Monte carlo deviation of estimates and "Coverage" the empirical coverage probability of posterior credible intervals.

5.6.1.1 Simulation study results for marker state probabilities

We also present results for marker state probabilities over time (probabilities of being event free and having "true" marker values in predefined intervals) from the SREM-CIF-1 and SREM-CIF-2 models in Tables 5.3-5.6. Our findings were similar, as both models, even when misspecified, yielded estimates with small biases and adequate coverage rates (91-97%).

Table 5.3: Simulation study results for marker state probabilities (%) for group 1 ($w = 1$) when the data have been generated by SREM-CIF-1 and there is no misclassification†.

State	Est	$t=0$	$t=2$	$t=4$	$t=6$	$t=8$	$t=10$	$t=0$	$t=2$	$t=4$	$t=6$	$t=8$	$t=10$
		Results from SREM-CIF-1						Results from SREM-CIF-2					
$\{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	True	12.431	0.183	0.037	0.020	0.005	0.001	12.431	0.183	0.037	0.020	0.005	0.001
$\{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	Median	12.414	0.187	0.040	0.023	0.007	0.002	12.407	0.189	0.044	0.026	0.008	0.002
$\{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	Bias	-0.017	0.003	0.004	0.003	0.002	0.001	-0.024	0.006	0.007	0.006	0.003	0.001
$\{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	Coverage	92.800	94.400	94.800	94.200	94.800	94.600	92.200	95.000	92.600	91.000	93.000	94.400
$\{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	True	16.038	1.007	0.307	0.190	0.103	0.050	16.038	1.007	0.307	0.190	0.103	0.050
$\{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	Median	16.025	1.017	0.322	0.204	0.112	0.058	16.026	1.006	0.326	0.209	0.119	0.063
$\{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	Bias	-0.013	0.009	0.015	0.014	0.010	0.007	-0.012	-0.002	0.019	0.019	0.016	0.012
$\{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	Coverage	92.200	94.600	95.600	94.400	94.800	95.600	92.600	94.000	96.000	93.200	93.800	93.600
$\{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	True	31.710	7.103	3.098	2.040	1.511	1.127	31.710	7.103	3.098	2.040	1.511	1.127
$\{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	Median	31.718	7.127	3.162	2.108	1.564	1.175	31.728	7.069	3.141	2.093	1.557	1.174
$\{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	Bias	0.008	0.024	0.064	0.068	0.053	0.048	0.019	-0.034	0.042	0.053	0.046	0.047
$\{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	Coverage	94.600	95.200	95.400	94.200	95.200	95.800	94.400	95.200	96.000	95.400	95.400	95.800
$\{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	True	12.103	6.468	3.468	2.364	1.909	1.569	12.103	6.468	3.468	2.364	1.909	1.569
$\{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	Median	12.104	6.476	3.499	2.401	1.943	1.601	12.107	6.451	3.474	2.378	1.919	1.580
$\{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	Bias	0.001	0.008	0.031	0.036	0.033	0.033	0.004	-0.017	0.007	0.014	0.010	0.012
$\{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	Coverage	93.400	95.400	95.000	95.600	95.400	96.200	93.400	95.000	95.600	97.000	96.000	96.600
$\{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	True	15.610	16.626	10.745	7.677	6.502	5.607	15.610	16.626	10.745	7.677	6.502	5.607
$\{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	Median	15.600	16.627	10.767	7.711	6.541	5.652	15.601	16.617	10.720	7.656	6.475	5.581
$\{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	Bias	-0.009	0.002	0.022	0.034	0.039	0.045	-0.009	-0.009	-0.025	-0.021	-0.027	-0.026
$\{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	Coverage	92.600	95.200	95.200	94.400	96.400	94.800	92.200	94.600	94.400	95.400	95.600	95.800
$\{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	True	9.229	24.230	20.064	15.723	13.847	12.389	9.229	24.230	20.064	15.723	13.847	12.389
$\{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	Median	9.216	24.211	20.007	15.657	13.804	12.365	9.211	24.261	19.993	15.611	13.727	12.266
$\{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	Bias	-0.013	-0.019	-0.057	-0.065	-0.043	-0.024	-0.018	0.031	-0.071	-0.112	-0.120	-0.122
$\{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	Coverage	92.600	93.400	94.400	94.200	95.200	94.600	91.800	94.600	93.800	93.800	93.600	93.800
$\{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	True	2.880	24.682	30.397	31.163	28.411	26.223	2.880	24.682	30.397	31.163	28.411	26.223
$\{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	Median	2.883	24.684	30.434	31.198	28.449	26.265	2.878	24.785	30.542	31.270	28.466	26.221
$\{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	Bias	0.003	0.002	0.038	0.034	0.038	0.042	-0.002	0.103	0.145	0.106	0.055	-0.002
$\{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	Coverage	93.000	96.000	94.000	92.600	92.400	92.800	93.000	94.200	93.400	92.800	92.000	92.400

† Results from 500 replications with each dataset including 1500 individuals. The true marker evolution was based on linear splines with knots at 1 and 5 years since baseline and it was correctly specified in the fitted SREM-CIF-1 and SREM-CIF-2 models. "True" denotes the true parameter values; "Median" the mean of posterior medians over the 500 replications; "Bias" the mean bias for posterior median estimates; and "Coverage" the empirical coverage probability of posterior credible intervals.

Table 5.4: Simulation study results for marker state probabilities (%) for group 0 ($w = 0$) when the data have been generated by SREM-CIF-1 and there is no misclassification†.

State	Est	$t=0$	$t=2$	$t=4$	$t=6$	$t=8$	$t=10$	$t=0$	$t=2$	$t=4$	$t=6$	$t=8$	$t=10$
		Results from SREM-CIF-1						Results from SREM-CIF-2					
$\{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	True	12.431	0.193	0.039	0.021	0.005	0.001	12.431	0.193	0.039	0.021	0.005	0.001
$\{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	Median	12.414	0.196	0.043	0.024	0.007	0.001	12.407	0.194	0.044	0.025	0.008	0.002
$\{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	Bias	-0.017	0.003	0.003	0.003	0.001	0.001	-0.024	0.001	0.005	0.004	0.002	0.001
$\{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	Coverage	92.800	95.600	95.000	93.800	95.800	94.800	92.200	95.800	94.800	93.800	94.800	95.600
$\{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	True	16.038	1.031	0.315	0.193	0.103	0.049	16.038	1.031	0.315	0.193	0.103	0.049
$\{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	Median	16.025	1.037	0.328	0.206	0.112	0.056	16.026	1.018	0.326	0.205	0.114	0.059
$\{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	Bias	-0.013	0.006	0.013	0.013	0.008	0.006	-0.012	-0.013	0.011	0.013	0.011	0.009
$\{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	Coverage	92.200	96.200	94.800	94.400	95.200	95.600	92.600	95.200	95.200	94.200	95.800	95.400
$\{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	True	31.710	7.136	3.098	2.016	1.474	1.081	31.710	7.136	3.098	2.016	1.474	1.081
$\{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	Median	31.718	7.150	3.153	2.075	1.519	1.120	31.728	7.066	3.108	2.045	1.504	1.120
$\{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	Bias	0.008	0.015	0.055	0.060	0.045	0.039	0.019	-0.070	0.009	0.030	0.030	0.039
$\{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	Coverage	94.600	94.800	94.200	94.800	96.400	95.600	94.400	94.800	94.400	95.200	96.200	95.800
$\{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	True	12.103	6.450	3.433	2.312	1.844	1.493	12.103	6.450	3.433	2.312	1.844	1.493
$\{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	Median	12.104	6.452	3.457	2.343	1.870	1.519	12.107	6.409	3.418	2.313	1.848	1.507
$\{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	Bias	0.001	0.002	0.025	0.030	0.026	0.026	0.004	-0.041	-0.015	0.001	0.005	0.014
$\{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	Coverage	93.400	95.000	95.000	96.000	95.600	95.400	93.400	93.800	95.600	95.600	95.800	96.000
$\{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	True	15.610	16.505	10.570	7.462	6.238	5.310	15.610	16.505	10.570	7.462	6.238	5.310
$\{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	Median	15.600	16.500	10.579	7.483	6.263	5.341	15.601	16.447	10.500	7.417	6.210	5.307
$\{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	Bias	-0.009	-0.006	0.010	0.022	0.025	0.032	-0.009	-0.058	-0.070	-0.045	-0.029	-0.003
$\{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	Coverage	92.600	93.200	95.200	95.600	95.400	95.400	92.200	93.400	94.600	95.600	94.800	95.400
$\{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	True	9.229	23.965	19.630	15.200	13.219	11.687	9.229	23.965	19.630	15.200	13.219	11.687
$\{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	Median	9.216	23.939	19.564	15.128	13.168	11.654	9.211	23.932	19.499	15.064	13.117	11.622
$\{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	Bias	-0.013	-0.026	-0.066	-0.072	-0.051	-0.033	-0.018	-0.034	-0.130	-0.136	-0.103	-0.065
$\{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	Coverage	92.600	94.600	93.600	95.200	94.200	95.000	91.800	95.200	93.800	94.200	94.400	95.200
$\{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	True	2.880	24.354	29.626	30.016	27.051	24.708	2.880	24.354	29.626	30.016	27.051	24.708
$\{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	Median	2.883	24.358	29.663	30.053	27.092	24.756	2.878	24.402	29.693	30.073	27.117	24.785
$\{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	Bias	0.003	0.004	0.037	0.036	0.042	0.048	-0.002	0.049	0.067	0.057	0.067	0.077
$\{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	Coverage	93.000	95.800	95.600	95.200	94.600	94.800	93.000	95.200	95.200	95.800	95.000	94.000

† Results from 500 replications with each dataset including 1500 individuals. The true marker evolution was based on linear splines with knots at 1 and 5 years since baseline and it was correctly specified in the fitted SREM-CIF-1 and SREM-CIF-2 models. "True" denotes the true parameter values; "Median" the mean of posterior medians over the 500 replications; "Bias" the mean bias for posterior median estimates; and "Coverage" the empirical coverage probability of posterior credible intervals.

Table 5.5: Simulation study results for marker state probabilities (%) for group 1 ($w = 1$) when the data have been generated by SREM-CIF-2 and there is no misclassification†.

State	Est	$t=0$	$t=2$	$t=4$	$t=6$	$t=8$	$t=10$	$t=0$	$t=2$	$t=4$	$t=6$	$t=8$	$t=10$
		Results from SREM-CIF-1						Results from SREM-CIF-2					
$\{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	True	12.431	0.198	0.048	0.031	0.014	0.005	12.431	0.198	0.048	0.031	0.014	0.005
$\{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	Median	12.530	0.203	0.050	0.033	0.015	0.006	12.526	0.203	0.052	0.035	0.016	0.007
$\{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	Bias	0.099	0.005	0.002	0.002	0.001	0.001	0.095	0.005	0.004	0.004	0.003	0.002
$\{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	Coverage	94.600	94.400	93.800	95.600	95.000	95.000	94.200	94.800	94.000	95.400	95.400	95.800
$\{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	True	16.038	1.015	0.333	0.223	0.145	0.092	16.038	1.015	0.333	0.223	0.145	0.092
$\{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	Median	16.026	1.040	0.347	0.237	0.153	0.098	16.028	1.026	0.347	0.238	0.157	0.102
$\{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	Bias	-0.011	0.025	0.014	0.014	0.008	0.005	-0.010	0.011	0.015	0.015	0.011	0.010
$\{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	Coverage	95.800	94.400	93.200	95.000	95.000	94.600	96.000	94.600	93.400	96.000	95.600	94.600
$\{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	True	31.710	6.990	3.111	2.110	1.643	1.305	31.710	6.990	3.111	2.110	1.643	1.305
$\{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	Median	31.617	7.075	3.192	2.193	1.712	1.364	31.624	7.018	3.166	2.173	1.697	1.354
$\{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	Bias	-0.093	0.085	0.081	0.083	0.069	0.060	-0.085	0.028	0.055	0.063	0.054	0.049
$\{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	Coverage	94.600	94.400	93.400	94.400	93.800	93.600	94.400	95.600	93.800	95.400	94.600	94.000
$\{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	True	12.103	6.336	3.415	2.364	1.953	1.650	12.103	6.336	3.415	2.364	1.953	1.650
$\{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	Median	12.060	6.367	3.462	2.416	2.006	1.703	12.062	6.343	3.438	2.394	1.982	1.680
$\{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	Bias	-0.043	0.031	0.047	0.052	0.053	0.052	-0.041	0.007	0.024	0.030	0.029	0.029
$\{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	Coverage	94.800	94.000	93.200	93.000	92.000	91.600	94.600	95.200	94.200	95.000	94.200	93.800
$\{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	True	15.610	16.278	10.503	7.565	6.483	5.674	15.610	16.278	10.503	7.565	6.483	5.674
$\{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	Median	15.566	16.280	10.549	7.628	6.565	5.769	15.567	16.268	10.506	7.579	6.506	5.703
$\{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	Bias	-0.044	0.003	0.046	0.064	0.082	0.095	-0.043	-0.009	0.003	0.015	0.022	0.030
$\{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	Coverage	94.400	95.400	94.200	94.000	93.000	92.200	94.600	94.400	94.000	93.800	94.200	94.000
$\{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	True	9.229	23.750	19.550	15.361	13.602	12.257	9.229	23.750	19.550	15.361	13.602	12.257
$\{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	Median	9.238	23.659	19.475	15.304	13.590	12.285	9.235	23.707	19.467	15.265	13.530	12.208
$\{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	Bias	0.009	-0.092	-0.076	-0.058	-0.012	0.028	0.006	-0.044	-0.084	-0.096	-0.072	-0.050
$\{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	Coverage	95.400	95.200	94.200	92.800	93.600	94.600	95.600	95.200	93.800	92.800	93.600	92.600
$\{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	True	2.880	24.253	29.643	30.360	27.724	25.659	2.880	24.253	29.643	30.360	27.724	25.659
$\{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	Median	2.920	24.146	29.538	30.270	27.682	25.679	2.917	24.247	29.645	30.348	27.724	25.681
$\{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	Bias	0.040	-0.107	-0.105	-0.090	-0.042	0.020	0.037	-0.006	0.002	-0.012	0.000	0.022
$\{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	Coverage	94.600	93.200	93.800	94.200	93.800	94.400	94.800	93.600	94.400	95.600	93.400	93.400

† Results from 500 replications with each dataset including 1500 individuals. The true marker evolution was based on linear splines with knots at 1 and 5 years since baseline and it was correctly specified in the fitted SREM-CIF-1 and SREM-CIF-2 models. "True" denotes the true parameter values; "Median" the mean of posterior medians over the 500 replications; "Bias" the mean bias for posterior median estimates; and "Coverage" the empirical coverage probability of posterior credible intervals.

Table 5.6: Simulation study results for marker state probabilities (%) for group 0 ($w = 0$) when the data have been generated by SREM-CIF-2 and there is no misclassification†.

State	Est	$t=0$	$t=2$	$t=4$	$t=6$	$t=8$	$t=10$	$t=0$	$t=2$	$t=4$	$t=6$	$t=8$	$t=10$
		Results from SREM-CIF-1						Results from SREM-CIF-2					
$\{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	True	12.431	0.203	0.049	0.031	0.014	0.005	12.431	0.203	0.049	0.031	0.014	0.005
$\{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	Median	12.530	0.210	0.052	0.034	0.015	0.006	12.526	0.207	0.052	0.035	0.016	0.007
$\{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	Bias	0.099	0.007	0.003	0.003	0.001	0.001	0.095	0.004	0.004	0.004	0.002	0.002
$\{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	Coverage	94.600	95.200	94.000	95.400	94.800	94.000	94.200	95.200	94.600	96.200	95.000	94.800
$\{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	True	16.038	1.028	0.335	0.223	0.145	0.092	16.038	1.028	0.335	0.223	0.145	0.092
$\{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	Median	16.026	1.056	0.352	0.238	0.154	0.097	16.028	1.036	0.348	0.236	0.154	0.099
$\{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	Bias	-0.011	0.028	0.017	0.016	0.009	0.006	-0.010	0.008	0.012	0.013	0.009	0.008
$\{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	Coverage	95.800	94.400	95.000	94.800	94.200	94.400	96.000	95.600	95.800	95.600	94.600	94.200
$\{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	True	31.710	7.000	3.099	2.087	1.617	1.279	31.710	7.000	3.099	2.087	1.617	1.279
$\{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	Median	31.617	7.089	3.184	2.168	1.679	1.326	31.624	7.011	3.139	2.135	1.656	1.313
$\{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	Bias	-0.093	0.089	0.086	0.081	0.062	0.047	-0.085	0.011	0.041	0.048	0.038	0.034
$\{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	Coverage	94.600	95.400	93.400	94.400	94.000	94.200	94.400	95.800	94.400	94.600	94.600	94.600
$\{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	True	12.103	6.308	3.380	2.324	1.910	1.607	12.103	6.308	3.380	2.324	1.910	1.607
$\{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	Median	12.060	6.343	3.427	2.370	1.950	1.640	12.062	6.303	3.390	2.341	1.926	1.622
$\{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	Bias	-0.043	0.035	0.048	0.046	0.040	0.033	-0.041	-0.005	0.011	0.017	0.015	0.016
$\{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	Coverage	94.800	95.400	94.200	93.800	94.400	94.600	94.600	95.400	95.000	94.400	95.400	95.400
$\{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	True	15.610	16.146	10.346	7.401	6.306	5.492	15.610	16.146	10.346	7.401	6.306	5.492
$\{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	Median	15.566	16.165	10.391	7.445	6.349	5.529	15.567	16.113	10.318	7.383	6.294	5.486
$\{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	Bias	-0.044	0.018	0.045	0.044	0.042	0.037	-0.043	-0.033	-0.028	-0.018	-0.013	-0.006
$\{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	Coverage	94.400	94.800	95.800	95.400	96.000	96.200	94.600	95.800	96.200	95.800	95.000	95.600
$\{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	True	9.229	23.477	19.167	14.953	13.157	11.794	9.229	23.477	19.167	14.953	13.157	11.794
$\{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	Median	9.238	23.414	19.097	14.866	13.080	11.724	9.235	23.405	19.039	14.809	13.032	11.689
$\{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	Bias	0.009	-0.063	-0.070	-0.088	-0.077	-0.071	0.006	-0.072	-0.128	-0.145	-0.125	-0.106
$\{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	Coverage	95.400	95.400	94.200	93.200	94.200	94.400	95.600	95.600	94.800	93.400	94.200	94.200
$\{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	True	2.880	23.914	28.940	29.411	26.674	24.551	2.880	23.914	28.940	29.411	26.674	24.551
$\{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	Median	2.920	23.841	28.856	29.293	26.550	24.442	2.917	23.883	28.885	29.321	26.589	24.485
$\{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	Bias	0.040	-0.073	-0.084	-0.118	-0.123	-0.109	0.037	-0.031	-0.055	-0.090	-0.085	-0.066
$\{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	Coverage	94.600	93.800	94.400	93.200	92.200	92.200	94.800	93.800	94.800	94.000	94.200	93.400

† Results from 500 replications with each dataset including 1500 individuals. The true marker evolution was based on linear splines with knots at 1 and 5 years since baseline and it was correctly specified in the fitted SREM-CIF-1 and SREM-CIF-2 models. "True" denotes the true parameter values; "Median" the mean of posterior medians over the 500 replications; "Bias" the mean bias for posterior median estimates; and "Coverage" the empirical coverage probability of posterior credible intervals.

5.6.1.2 Simulation study results for transition marker state probabilities by baseline marker state

In this subsection, we provide results for transition marker probabilities by baseline marker state in Tables 5.7-5.10. In nearly all cases, both models performed satisfactorily with small biases and coverage rates close to the nominal level, except for some cases where the true transition probability was very low and the fitted model was misspecified. For example, when the data were generated from the SREM-CIF-1 model and the SREM-CIF-2 model was fitted, the coverage probability for the "$m_i(0) < \sqrt{50} \rightarrow \{m_i(6) < \sqrt{50}\} \cap (T_i^\star > 6)$" transition was equal to 85.4% (Table 5.7).

Table 5.7: Simulation study results for transition marker state probabilities (%) by baseline marker state for group 1 ($w = 1$) when the data have been generated by SREM-CIF-1 and there is no misclassification†.

Transition	$t=2$	$t=4$	$t=6$	$t=8$	$t=10$	$t=2$	$t=4$	$t=6$	$t=8$	$t=10$
	Results from SREM-CIF-1					Results from SREM-CIF-2				
$m_i(0) < \sqrt{50} \rightarrow \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	1.31/1.33/93.6	0.15/0.17/93.8	0.03/0.04/93.2	<0.01/0.01/94.8	<0.01/<0.01/95.4	1.31/1.36/93.8	0.15/0.2/89.4	0.03/0.06/85.4	<0.01/0.01/87.4	<0.01/<0.01/91.4
$m_i(0) < \sqrt{50} \rightarrow \{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	5.86/5.91/92	1.09/1.14/94.6	0.36/0.4/94.4	0.13/0.16/95.2	0.03/0.05/95.4	5.86/5.86/92.8	1.09/1.2/91.8	0.36/0.46/88	0.13/0.21/86	0.03/0.08/88.2
$m_i(0) < \sqrt{50} \rightarrow \{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	24.64/24.72/93.6	7.77/7.95/95	3.44/3.59/94.6	2.25/2.37/95.4	1.42/1.53/95.4	24.64/24.44/93.6	7.77/7.99/95.2	3.44/3.72/91.2	2.25/2.55/90.6	1.42/1.72/90.4
$m_i(0) < \sqrt{50} \rightarrow \{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	12.55/12.52/94	6.45/6.5/94.6	3.4/3.46/95	2.57/2.63/95.2	1.95/2/95	12.55/12.39/92.4	6.45/6.47/96	3.4/3.5/94.8	2.57/2.7/93.8	1.95/2.09/93.8
$m_i(0) < \sqrt{50} \rightarrow \{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	15.7/15.58/93.2	14.04/14.01/95.4	8.96/8.99/95	7.26/7.3/94.6	5.96/6.01/96	15.7/15.43/92.2	14.04/13.9/95	8.96/9/94.6	7.26/7.36/94.2	5.96/6.1/95.2
$m_i(0) < \sqrt{50} \rightarrow \{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	6.3/6.21/93.2	14.75/14.57/94.2	13.06/12.92/93.8	11.18/11.07/93.8	9.71/9.62/94.2	6.3/6.16/93	14.75/14.45/92.4	13.06/12.87/94.6	11.18/11.06/93.6	9.71/9.64/94.4
$m_i(0) < \sqrt{50} \rightarrow \{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	0.66/0.66/93.2	7.56/7.43/93	12.77/12.58/93.6	11.4/11.23/94.2	10.31/10.15/94	0.66/0.65/93.2	7.56/7.37/92.8	12.77/12.51/94	11.4/11.18/93.8	10.31/10.1/94.6
$\sqrt{50} \leq m_i(0) < \sqrt{100} \rightarrow \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	0.11/0.11/93.6	0.06/0.07/94.2	0.03/0.04/94.2	0.01/0.01/95	<0.01/<0.01/94.8	0.11/0.11/93	0.06/0.07/94.8	0.03/0.04/91.8	0.01/0.01/91.6	<0.01/<0.01/91.8
$\sqrt{50} \leq m_i(0) < \sqrt{100} \rightarrow \{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	1.27/1.28/93.2	0.52/0.55/94.6	0.31/0.33/94.2	0.17/0.18/94.2	0.07/0.08/95	1.27/1.27/92.8	0.52/0.54/94.6	0.31/0.34/94.4	0.17/0.19/92.6	0.07/0.1/92.2
$\sqrt{50} \leq m_i(0) < \sqrt{100} \rightarrow \{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	13.93/14.02/93.4	5.2/5.32/94.6	3.03/3.14/94.2	2.23/2.31/93.8	1.64/1.71/94.8	13.93/13.92/93.2	5.2/5.26/94.6	3.03/3.1/95.2	2.23/2.3/94	1.64/1.72/95
$\sqrt{50} \leq m_i(0) < \sqrt{100} \rightarrow \{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	12.83/12.89/93	5.45/5.52/94.4	3.24/3.3/95	2.6/2.65/95.4	2.11/2.17/95.2	12.83/12.84/93.4	5.45/5.47/96.2	3.24/3.27/96.8	2.6/2.62/96.2	2.11/2.14/95.8
$\sqrt{50} \leq m_i(0) < \sqrt{100} \rightarrow \{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	25.61/25.63/92.6	14.79/14.88/93.2	9.5/9.58/94.8	7.98/8.06/94.6	6.82/6.9/94	25.61/25.55/92.6	14.79/14.78/94.2	9.5/9.5/95	7.98/7.98/94.8	6.82/6.83/94.8
$\sqrt{50} \leq m_i(0) < \sqrt{100} \rightarrow \{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	18.56/18.43/93.2	20.82/20.76/93.4	16.27/16.22/93.8	14.21/14.18/94.8	12.61/12.6/94.2	18.56/18.38/92.8	20.82/20.67/94.8	16.27/16.13/94.8	14.21/14.1/94.6	12.61/12.51/92.8
$\sqrt{50} \leq m_i(0) < \sqrt{100} \rightarrow \{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	3.83/3.78/92.4	15.7/15.56/93.2	20.71/20.56/94	18.71/18.57/93	17.11/16.99/92.8	3.83/3.77/92.6	15.7/15.53/92.4	20.71/20.52/93.6	18.71/18.53/93.6	17.11/16.94/92.2
$\sqrt{100} \leq m_i(0) < \sqrt{200} \rightarrow \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	0.01/0.01/94.4	0.02/0.02/93.8	0.02/0.02/94.8	0.01/0.01/94.6	<0.01/<0.01/94.8	0.01/0.01/93	0.02/0.02/93.8	0.02/0.03/95.2	0.01/0.01/94	<0.01/<0.01/93.6
$\sqrt{100} \leq m_i(0) < \sqrt{200} \rightarrow \{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	0.22/0.23/94.2	0.22/0.23/94	0.2/0.21/95	0.12/0.13/94.6	0.07/0.07/96	0.22/0.22/93.2	0.22/0.22/94.6	0.2/0.21/95.2	0.12/0.13/95.6	0.07/0.07/94.4
$\sqrt{100} \leq m_i(0) < \sqrt{200} \rightarrow \{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	4.96/4.96/94.2	2.93/2.98/94.2	2.19/2.26/94.8	1.68/1.73/95	1.3/1.35/95.8	4.96/4.95/94.2	2.93/2.94/94.6	2.19/2.21/95	1.68/1.69/96.4	1.3/1.31/96.8
$\sqrt{100} \leq m_i(0) < \sqrt{200} \rightarrow \{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	7.19/7.2/93.6	3.73/3.76/94.6	2.6/2.64/95.6	2.13/2.17/95.8	1.78/1.81/96	7.19/7.19/93.4	3.73/3.73/94.8	2.6/2.6/96.6	2.13/2.12/96	1.78/1.77/96.6
$\sqrt{100} \leq m_i(0) < \sqrt{200} \rightarrow \{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	22.83/22.9/92.6	12.37/12.43/94.2	8.53/8.59/94.8	7.27/7.33/95.6	6.31/6.37/95.2	22.83/22.92/93.4	12.37/12.38/94	8.53/8.51/94.8	7.27/7.23/95.2	6.31/6.27/95.2
$\sqrt{100} \leq m_i(0) < \sqrt{200} \rightarrow \{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	31.24/31.27/92.6	23.02/23.03/92	17.15/17.13/93.6	15.12/15.13/94.6	13.55/13.57/94.2	31.24/31.32/91.8	23.02/23.02/93	17.15/17.08/93.6	15.12/15.04/94	13.55/13.45/93.6
$\sqrt{100} \leq m_i(0) < \sqrt{200} \rightarrow \{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	14.48/14.4/91.8	26.38/26.34/93.6	28.92/28.9/92.8	26.31/26.29/92.4	24.23/24.22/90.6	14.48/14.42/92.2	26.38/26.4/93.6	28.92/28.94/92.4	26.31/26.29/92	24.23/24.17/90.6
$\sqrt{200} \leq m_i(0) < \sqrt{250} \rightarrow \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	<0.01/<0.01/93.8	0.01/0.01/93.2	0.01/0.01/94	0.01/0.01/94.8	<0.01/<0.01/94.8	<0.01/<0.01/93.2	0.01/0.01/92.8	0.01/0.01/94.4	0.01/0.01/94.8	<0.01/<0.01/95
$\sqrt{200} \leq m_i(0) < \sqrt{250} \rightarrow \{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	0.03/0.03/94	0.09/0.09/93.4	0.12/0.13/93.4	0.08/0.08/95	0.05/0.05/95.6	0.03/0.03/93.6	0.09/0.09/93	0.12/0.12/94.8	0.08/0.08/96	0.05/0.05/95.4
$\sqrt{200} \leq m_i(0) < \sqrt{250} \rightarrow \{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	1.32/1.32/93.8	1.56/1.58/94.2	1.53/1.57/95	1.2/1.23/95.8	0.96/0.98/95.2	1.32/1.32/93.6	1.56/1.56/94	1.53/1.53/94.4	1.2/1.19/94.6	0.96/0.94/95.6
$\sqrt{200} \leq m_i(0) < \sqrt{250} \rightarrow \{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	2.94/2.93/94.2	2.37/2.38/94.6	2/2.02/95.6	1.66/1.68/95	1.4/1.42/95	2.94/2.93/93.8	2.37/2.36/93.8	2/1.99/94.8	1.66/1.64/95	1.4/1.38/94.4
$\sqrt{200} \leq m_i(0) < \sqrt{250} \rightarrow \{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	14.64/14.61/94.4	9.45/9.45/94.6	7.26/7.27/94.6	6.24/6.26/95.4	5.46/5.49/95	14.64/14.66/94.8	9.45/9.44/94.6	7.26/7.21/95	6.24/6.17/95	5.46/5.39/94.8
$\sqrt{200} \leq m_i(0) < \sqrt{250} \rightarrow \{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	34.65/34.72/92.4	22.37/22.37/92.6	16.78/16.74/93.6	14.88/14.86/94.6	13.4/13.4/95	34.65/34.85/92.2	22.37/22.4/93.2	16.78/16.71/94.4	14.88/14.77/94	13.4/13.28/93.6
$\sqrt{200} \leq m_i(0) < \sqrt{250} \rightarrow \{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	30.23/30.24/94	36.64/36.75/94.6	36.1/36.2/93.8	32.96/33.07/92.6	30.46/30.57/92.2	30.23/30.34/93.8	36.64/36.9/93.4	36.1/36.32/92.8	32.96/33.11/93	30.46/30.54/92.2
$\sqrt{250} \leq m_i(0) < \sqrt{350} \rightarrow \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	<0.01/<0.01/93.6	<0.01/<0.01/93	0.01/0.01/94.4	<0.01/<0.01/95.4	<0.01/<0.01/95.4	<0.01/<0.01/93.4	<0.01/<0.01/92.2	0.01/0.01/94.2	<0.01/<0.01/95.4	<0.01/<0.01/94.8
$\sqrt{250} \leq m_i(0) < \sqrt{350} \rightarrow \{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	0.01/0.01/93.8	0.04/0.04/92.4	0.08/0.08/93.6	0.05/0.06/94.6	0.04/0.04/95.6	0.01/0.01/93.6	0.04/0.04/92.8	0.08/0.08/93.6	0.05/0.05/94.2	0.04/0.03/94.8
$\sqrt{250} \leq m_i(0) < \sqrt{350} \rightarrow \{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	0.43/0.43/93.6	0.91/0.92/93.4	1.12/1.14/95.2	0.88/0.9/94.4	0.71/0.73/95	0.43/0.43/93.8	0.91/0.91/93.6	1.12/1.11/94.6	0.88/0.87/94	0.71/0.69/93.8
$\sqrt{250} \leq m_i(0) < \sqrt{350} \rightarrow \{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	1.22/1.21/93.6	1.56/1.56/94.2	1.56/1.57/95	1.3/1.31/94.2	1.11/1.12/94.4	1.22/1.22/93.8	1.56/1.55/94.2	1.56/1.55/95	1.3/1.28/94	1.11/1.08/93.6
$\sqrt{250} \leq m_i(0) < \sqrt{350} \rightarrow \{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	8.27/8.22/93.4	7.11/7.08/95.2	6.12/6.11/95.4	5.29/5.29/94.4	4.65/4.66/94.6	8.27/8.25/94.2	7.11/7.08/94.4	6.12/6.07/94.4	5.29/5.2/94.4	4.65/4.56/93.2
$\sqrt{250} \leq m_i(0) < \sqrt{350} \rightarrow \{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	29.51/29.47/93	20.18/20.11/94.4	15.77/15.69/94	14.02/13.96/95	12.67/12.63/94.6	29.51/29.61/93	20.18/20.16/94.8	15.77/15.66/94.2	14.02/13.88/95	12.67/12.5/93.6
$\sqrt{250} \leq m_i(0) < \sqrt{350} \rightarrow \{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	46.01/46.12/94	44.93/45.14/93.6	41.65/41.85/94.2	38.13/38.33/94.2	35.34/35.52/94	46.01/46.35/94.4	44.93/45.38/92.4	41.65/42.03/93.8	38.13/38.4/94	35.34/35.49/93.8
$\sqrt{350} \leq m_i(0) < \sqrt{500} \rightarrow \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	<0.01/<0.01/93.6	<0.01/<0.01/92.4	<0.01/<0.01/94	<0.01/<0.01/94.2	<0.01/<0.01/94.8	<0.01/<0.01/93.4	<0.01/<0.01/91.8	<0.01/<0.01/93.6	<0.01/<0.01/93.2	<0.01/<0.01/94.2
$\sqrt{350} \leq m_i(0) < \sqrt{500} \rightarrow \{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	<0.01/<0.01/93.8	0.01/0.01/92.4	0.04/0.04/94	0.03/0.03/94.6	0.02/0.02/94.8	<0.01/<0.01/94	0.01/0.01/92.8	0.04/0.04/94.4	0.03/0.03/92.8	0.02/0.02/92.6
$\sqrt{350} \leq m_i(0) < \sqrt{500} \rightarrow \{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	0.06/0.07/93.8	0.38/0.39/93.6	0.67/0.68/94.6	0.54/0.55/94.8	0.44/0.45/95	0.06/0.07/94	0.38/0.38/93.4	0.67/0.66/94.4	0.54/0.52/93.4	0.44/0.42/93.8
$\sqrt{350} \leq m_i(0) < \sqrt{500} \rightarrow \{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	0.26/0.26/94	0.78/0.77/94.6	1.04/1.04/95.6	0.87/0.87/94	0.75/0.75/94.6	0.26/0.26/94.2	0.78/0.77/94.2	1.04/1.02/94.6	0.87/0.85/93.4	0.75/0.72/93.6
$\sqrt{350} \leq m_i(0) < \sqrt{500} \rightarrow \{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	2.72/2.69/94	4.29/4.25/94.2	4.56/4.52/94.8	3.95/3.93/94	3.49/3.48/94.4	2.72/2.71/94.2	4.29/4.25/94.4	4.56/4.48/93.6	3.95/3.86/93.8	3.49/3.4/93.4
$\sqrt{350} \leq m_i(0) < \sqrt{500} \rightarrow \{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	17.44/17.31/94.6	15.79/15.63/95.2	13.69/13.54/94	12.21/12.09/94.2	11.07/10.97/93.8	17.44/17.39/94.6	15.79/15.68/95.4	13.69/13.52/93.8	12.21/12.01/94.6	11.07/10.85/93.6
$\sqrt{350} \leq m_i(0) < \sqrt{500} \rightarrow \{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	66.74/66.91/95.6	56.02/56.32/94.4	49.19/49.49/94	45.17/45.46/94.8	41.98/42.25/95	66.74/67.31/94.6	56.02/56.68/92.6	49.19/49.73/94.4	45.17/45.56/94.4	41.98/42.21/95.4
$\sqrt{500} \leq m_i(0) \rightarrow \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	<0.01/<0.01/93.8	<0.01/<0.01/93.4	<0.01/<0.01/94.2	<0.01/<0.01/93.2	<0.01/<0.01/94	<0.01/<0.01/93.6	<0.01/<0.01/93.2	<0.01/<0.01/93.8	<0.01/<0.01/93.2	<0.01/<0.01/92.6
$\sqrt{500} \leq m_i(0) \rightarrow \{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	<0.01/<0.01/94	<0.01/<0.01/93	0.02/0.02/95.4	0.01/0.01/95	0.01/0.01/94.4	<0.01/<0.01/93.6	<0.01/<0.01/93.4	0.02/0.02/94.2	0.01/0.01/93.2	0.01/0.01/93
$\sqrt{500} \leq m_i(0) \rightarrow \{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	<0.01/<0.01/94	0.11/0.11/93.6	0.31/0.32/95.2	0.25/0.26/94.4	0.21/0.22/94.8	<0.01/<0.01/94.2	0.11/0.11/94.4	0.31/0.31/94.4	0.25/0.25/94	0.21/0.2/93.2
$\sqrt{500} \leq m_i(0) \rightarrow \{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	0.02/0.02/94	0.27/0.27/94.2	0.55/0.55/94.8	0.47/0.47/94.8	0.4/0.41/94.6	0.02/0.02/94.4	0.27/0.27/94.6	0.55/0.54/94	0.47/0.45/93.8	0.4/0.39/94.4
$\sqrt{500} \leq m_i(0) \rightarrow \{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	0.4/0.41/95	1.88/1.85/94.6	2.81/2.77/94.6	2.45/2.42/95	2.17/2.15/95.6	0.4/0.41/94.8	1.88/1.85/94.6	2.81/2.75/94.4	2.45/2.37/94	2.17/2.09/94.4
$\sqrt{500} \leq m_i(0) \rightarrow \{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	5.36/5.3/93.6	9.59/9.4/95.2	10.33/10.13/95	9.24/9.07/94.6	8.4/8.26/94.8	5.36/5.33/94.2	9.59/9.43/94.8	10.33/10.11/94.6	9.24/9/94.4	8.4/8.15/94.4
$\sqrt{500} \leq m_i(0) \rightarrow \{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	83.1/83.14/94.4	67.94/68.2/94.6	58.09/58.4/94	53.53/53.85/94.4	49.9/50.21/94.8	83.1/83.69/93.8	67.94/68.7/93	58.09/58.73/93.4	53.53/53.97/94.4	49.9/50.15/94.8

† Results from 500 replications with each dataset including 1500 individuals. The true marker evolution was based on linear splines with knots at 1 and 5 years since baseline and it was correctly specified in the fitted SREM-CIF-1 and SREM-CIF-2 models. The true parameter values, the mean of posterior medians over the 500 replications and the empirical coverage probability of posterior credible intervals are presented, respectively.

Table 5.8: Simulation study results for transition marker state probabilities (%) by baseline marker state for group 0 ($w = 0$) when the data have been generated by SREM-CIF-1 and there is no misclassification†.

transition	$t=2$	$t=4$	$t=6$	$t=8$	$t=10$	$t=2$	$t=4$	$t=6$	$t=8$	$t=10$
	Results from SREM-CIF-1					Results from SREM-CIF-2				
$m_i(0) < \sqrt{50} \to \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	1.38/1.4/95	0.17/0.18/96	0.03/0.04/94.8	<0.01/0.01/96.2	<0.01/<0.01/95.6	1.38/1.39/95.2	0.17/0.2/93.2	0.03/0.06/91.4	<0.01/0.01/93.6	<0.01/<0.01/95.4
$m_i(0) < \sqrt{50} \to \{\sqrt{50} \le m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	6.02/6.06/94.4	1.14/1.19/95.6	0.38/0.41/95.6	0.14/0.16/96	0.03/0.05/96.2	6.02/5.95/94.4	1.14/1.2/95.6	0.38/0.45/93.4	0.14/0.2/92.8	0.03/0.07/93
$m_i(0) < \sqrt{50} \to \{\sqrt{100} \le m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	24.91/24.95/94.4	7.9/8.05/94.4	3.48/3.61/95.8	2.25/2.34/95.8	1.39/1.47/95.6	24.91/24.54/91.8	7.9/7.97/95.4	3.48/3.66/95.2	2.25/2.47/95	1.39/1.63/93.8
$m_i(0) < \sqrt{50} \to \{\sqrt{200} \le m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	12.6/12.54/94.8	6.47/6.51/95	3.39/3.44/96.2	2.53/2.57/96.6	1.88/1.92/96	12.6/12.37/91	6.47/6.42/95.8	3.39/3.43/96.8	2.53/2.61/95.8	1.88/2/94.8
$m_i(0) < \sqrt{50} \to \{\sqrt{250} \le m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	15.69/15.55/93.8	13.99/13.93/96.4	8.85/8.85/95.8	7.07/7.09/95.6	5.72/5.74/95.8	15.69/15.36/92.6	13.99/13.74/94.4	8.85/8.8/96.6	7.07/7.12/96.6	5.72/5.84/95.8
$m_i(0) < \sqrt{50} \to \{\sqrt{350} \le m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	6.27/6.18/93.4	14.6/14.39/94.6	12.8/12.64/95.4	10.81/10.68/95	9.27/9.16/96	6.27/6.11/93.4	14.6/14.22/93.8	12.8/12.54/95.6	10.81/10.66/95.8	9.27/9.21/96.2
$m_i(0) < \sqrt{50} \to \{\sqrt{500} \le m_i(t)\} \cap T_i^\star > t$	0.66/0.65/92.8	7.44/7.3/93.8	12.44/12.23/95	10.96/10.78/93.4	9.8/9.64/94	0.66/0.64/92.6	7.44/7.23/93.2	12.44/12.14/94.8	10.96/10.74/94.2	9.8/9.63/93.8
$\sqrt{50} \le m_i(0) < \sqrt{100} \to \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	0.11/0.11/93.6	0.06/0.07/94.2	0.04/0.04/95.6	0.01/0.01/96.6	<0.01/<0.01/94	0.11/0.11/92.6	0.06/0.07/94.2	0.04/0.04/94.4	0.01/0.01/95.2	<0.01/<0.01/95.2
$\sqrt{50} \le m_i(0) < \sqrt{100} \to \{\sqrt{50} \le m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	1.29/1.29/93	0.53/0.55/94	0.32/0.34/95.2	0.17/0.18/96	0.07/0.08/95.6	1.29/1.27/93	0.53/0.54/94.2	0.32/0.33/95.4	0.17/0.19/96.4	0.07/0.09/95
$\sqrt{50} \le m_i(0) < \sqrt{100} \to \{\sqrt{100} \le m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	13.95/14.03/94.4	5.19/5.3/95.2	3/3.09/96.6	2.18/2.25/96.4	1.58/1.63/95.6	13.95/13.88/94.8	5.19/5.21/94.8	3/3.04/96.8	2.18/2.22/96.8	1.58/1.64/96.4
$\sqrt{50} \le m_i(0) < \sqrt{100} \to \{\sqrt{200} \le m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	12.8/12.85/92	5.4/5.46/94.4	3.18/3.23/95.6	2.52/2.56/96.2	2.02/2.06/95.8	12.8/12.76/92.8	5.4/5.39/94.8	3.18/3.19/96.8	2.52/2.53/96.8	2.02/2.05/96.2
$\sqrt{50} \le m_i(0) < \sqrt{100} \to \{\sqrt{250} \le m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	25.47/25.47/93.8	14.59/14.65/93.2	9.27/9.33/95.4	7.68/7.74/96.2	6.48/6.54/95	25.47/25.33/94.4	14.59/14.5/93.8	9.27/9.23/95.6	7.68/7.68/96.4	6.48/6.51/96
$\sqrt{50} \le m_i(0) < \sqrt{100} \to \{\sqrt{350} \le m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	18.41/18.28/92.8	20.45/20.36/94	15.79/15.73/95.4	13.62/13.57/95.2	11.93/11.91/94.2	18.41/18.18/92.6	20.45/20.22/94	15.79/15.62/95.2	13.62/13.51/95.4	11.93/11.89/95.2
$\sqrt{50} \le m_i(0) < \sqrt{100} \to \{\sqrt{500} \le m_i(t)\} \cap T_i^\star > t$	3.79/3.74/92	15.36/15.22/93	20.02/19.87/94	17.87/17.73/94	16.17/16.05/94.2	3.79/3.72/92.2	15.36/15.15/93.4	20.02/19.81/94.2	17.87/17.72/93.8	16.17/16.06/94.6
$\sqrt{100} \le m_i(0) < \sqrt{200} \to \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	0.01/0.01/94	0.02/0.02/93.2	0.02/0.03/95.2	0.01/0.01/95.8	<0.01/<0.01/94.4	0.01/0.01/93.2	0.02/0.02/93.8	0.02/0.02/94.6	0.01/0.01/95	<0.01/<0.01/95.6
$\sqrt{100} \le m_i(0) < \sqrt{200} \to \{\sqrt{50} \le m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	0.23/0.23/94.2	0.22/0.23/93.8	0.2/0.21/94.4	0.12/0.13/96.4	0.07/0.07/95.6	0.23/0.22/93.2	0.22/0.22/94.4	0.2/0.2/94.6	0.12/0.12/96.4	0.07/0.07/96.2
$\sqrt{100} \le m_i(0) < \sqrt{200} \to \{\sqrt{100} \le m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	4.95/4.95/94.2	2.9/2.95/94	2.15/2.21/95.2	1.63/1.67/96	1.25/1.28/96	4.95/4.92/94	2.9/2.9/94.6	2.15/2.16/95.6	1.63/1.63/96	1.25/1.25/96.8
$\sqrt{100} \le m_i(0) < \sqrt{200} \to \{\sqrt{200} \le m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	7.14/7.15/93.4	3.67/3.7/94.8	2.53/2.57/96.8	2.05/2.08/97	1.69/1.72/96.2	7.14/7.13/94.2	3.67/3.66/95.8	2.53/2.53/96.8	2.05/2.04/97.2	1.69/1.69/96.8
$\sqrt{100} \le m_i(0) < \sqrt{200} \to \{\sqrt{250} \le m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	22.64/22.7/93.8	12.14/12.18/93.2	8.27/8.31/95	6.96/7/95.4	5.96/6.01/95	22.64/22.66/93.8	12.14/12.11/93.4	8.27/8.23/94.2	6.96/6.93/95.6	5.96/5.96/96.2
$\sqrt{100} \le m_i(0) < \sqrt{200} \to \{\sqrt{350} \le m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	30.91/30.93/92.6	22.51/22.51/94	16.56/16.54/94.4	14.43/14.42/94.2	12.77/12.78/94	30.91/30.9/92.8	22.51/22.44/94	16.56/16.47/94.4	14.43/14.36/94.2	12.77/12.74/94.6
$\sqrt{100} \le m_i(0) < \sqrt{200} \to \{\sqrt{500} \le m_i(t)\} \cap T_i^\star > t$	14.31/14.23/92.6	25.73/25.69/94	27.87/27.85/95.2	25.05/25.04/93.6	22.83/22.82/93.6	14.31/14.22/92.6	25.73/25.68/94	27.87/27.85/95.6	25.05/25.06/94	22.83/22.86/93.4
$\sqrt{200} \le m_i(0) < \sqrt{250} \to \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	<0.01/<0.01/93.6	0.01/0.01/91.6	0.01/0.01/93.8	0.01/0.01/94.8	<0.01/<0.01/94.8	<0.01/<0.01/93	0.01/0.01/93	0.01/0.01/94.2	0.01/0.01/95.2	<0.01/<0.01/95.4
$\sqrt{200} \le m_i(0) < \sqrt{250} \to \{\sqrt{50} \le m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	0.03/0.03/94	0.09/0.09/92.8	0.12/0.13/93.8	0.08/0.08/94.6	0.05/0.05/95.6	0.03/0.03/93.6	0.09/0.09/93.2	0.12/0.12/94.6	0.08/0.08/95	0.05/0.05/96.2
$\sqrt{200} \le m_i(0) < \sqrt{250} \to \{\sqrt{100} \le m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	1.32/1.31/93.8	1.54/1.55/93.2	1.5/1.53/94.4	1.16/1.18/95	0.91/0.93/95.4	1.32/1.31/93.4	1.54/1.53/92.8	1.5/1.49/94.6	1.16/1.14/95.4	0.91/0.9/96.4
$\sqrt{200} \le m_i(0) < \sqrt{250} \to \{\sqrt{200} \le m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	2.92/2.9/93.8	2.33/2.33/94.6	1.94/1.96/94.6	1.59/1.6/95.8	1.33/1.34/95.8	2.92/2.9/93.8	2.33/2.31/95.6	1.94/1.92/94.8	1.59/1.57/96.6	1.33/1.31/96.2
$\sqrt{200} \le m_i(0) < \sqrt{250} \to \{\sqrt{250} \le m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	14.49/14.46/92.8	9.24/9.24/94.4	7.01/7.02/95	5.95/5.97/95.8	5.15/5.17/95.8	14.49/14.47/93.2	9.24/9.21/95	7.01/6.96/94.6	5.95/5.9/94.8	5.15/5.11/95.4
$\sqrt{200} \le m_i(0) < \sqrt{250} \to \{\sqrt{350} \le m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	34.23/34.3/94.6	21.83/21.82/94.2	16.16/16.12/94.2	14.16/14.13/94.4	12.61/12.6/94.2	34.23/34.35/94.8	21.83/21.8/94.2	16.16/16.08/94.6	14.16/14.08/95	12.61/12.55/95
$\sqrt{200} \le m_i(0) < \sqrt{250} \to \{\sqrt{500} \le m_i(t)\} \cap T_i^\star > t$	29.84/29.84/93.8	35.69/35.79/94.8	34.72/34.84/95.4	31.35/31.46/94.2	28.67/28.79/94.4	29.84/29.88/93.4	35.69/35.85/95.4	34.72/34.89/95.6	31.35/31.51/94.8	28.67/28.84/95.2
$\sqrt{250} \le m_i(0) < \sqrt{350} \to \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	<0.01/<0.01/93.4	<0.01/<0.01/92.4	0.01/0.01/93.8	<0.01/<0.01/94.8	<0.01/<0.01/95.6	<0.01/<0.01/93.4	<0.01/<0.01/92.4	0.01/0.01/94	<0.01/<0.01/95.4	<0.01/<0.01/96
$\sqrt{250} \le m_i(0) < \sqrt{350} \to \{\sqrt{50} \le m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	0.01/0.01/93.6	0.04/0.04/93	0.08/0.08/94	0.05/0.05/94.6	0.03/0.04/95.6	0.01/0.01/93.4	0.04/0.04/92.8	0.08/0.08/93.8	0.05/0.05/94.6	0.03/0.03/96
$\sqrt{250} \le m_i(0) < \sqrt{350} \to \{\sqrt{100} \le m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	0.43/0.43/93.4	0.9/0.9/93.6	1.08/1.1/94.4	0.85/0.86/94.8	0.68/0.69/95	0.43/0.43/93.2	0.9/0.89/93.8	1.08/1.08/94	0.85/0.83/94.6	0.68/0.66/95.4
$\sqrt{250} \le m_i(0) < \sqrt{350} \to \{\sqrt{200} \le m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	1.21/1.2/94	1.53/1.53/94.2	1.51/1.52/95.4	1.25/1.25/95.8	1.05/1.06/96	1.21/1.2/94	1.53/1.52/94.4	1.51/1.49/94.6	1.25/1.22/95.2	1.05/1.03/94.8
$\sqrt{250} \le m_i(0) < \sqrt{350} \to \{\sqrt{250} \le m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	8.18/8.12/94.2	6.94/6.92/94.4	5.91/5.9/94.8	5.04/5.03/95.6	4.38/4.38/95	8.18/8.13/94.2	6.94/6.89/94.6	5.91/5.84/94.4	5.04/4.97/94.8	4.38/4.32/95.4
$\sqrt{250} \le m_i(0) < \sqrt{350} \to \{\sqrt{350} \le m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	29.14/29.1/93	19.67/19.6/94.4	15.17/15.09/94.6	13.33/13.27/93.8	11.91/11.87/93.8	29.14/29.17/92.8	19.67/19.6/94.4	15.17/15.05/94.2	13.33/13.21/93.4	11.91/11.81/94
$\sqrt{250} \le m_i(0) < \sqrt{350} \to \{\sqrt{500} \le m_i(t)\} \cap T_i^\star > t$	45.38/45.5/94.2	43.73/43.94/94.2	40.04/40.25/94.8	36.24/36.45/95	33.25/33.44/95	45.38/45.62/94.6	43.73/44.07/94.4	40.04/40.35/95.2	36.24/36.52/95	33.25/33.49/94.8
$\sqrt{350} \le m_i(0) < \sqrt{500} \to \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	<0.01/<0.01/93.8	<0.01/<0.01/92.2	<0.01/<0.01/94.2	<0.01/<0.01/94.2	<0.01/<0.01/95.6	<0.01/<0.01/93.6	<0.01/<0.01/92.2	<0.01/<0.01/93.8	<0.01/<0.01/93.4	<0.01/<0.01/95
$\sqrt{350} \le m_i(0) < \sqrt{500} \to \{\sqrt{50} \le m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	<0.01/<0.01/94	0.01/0.01/92.8	0.04/0.04/94.8	0.03/0.03/94	0.02/0.02/94.8	<0.01/<0.01/94.2	0.01/0.01/92.6	0.04/0.04/94.2	0.03/0.03/94	0.02/0.02/94.8
$\sqrt{350} \le m_i(0) < \sqrt{500} \to \{\sqrt{100} \le m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	0.06/0.06/93.8	0.38/0.38/93	0.65/0.66/95	0.52/0.52/94.8	0.42/0.42/94.4	0.06/0.06/93.4	0.38/0.38/93.4	0.65/0.64/94.6	0.52/0.5/94.8	0.42/0.4/94.4
$\sqrt{350} \le m_i(0) < \sqrt{500} \to \{\sqrt{200} \le m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	0.25/0.25/94	0.76/0.76/94	1/1/95.2	0.83/0.83/95.4	0.71/0.71/95	0.25/0.25/94.2	0.76/0.75/94	1/0.99/94.6	0.83/0.81/94.6	0.71/0.69/94.6
$\sqrt{350} \le m_i(0) < \sqrt{500} \to \{\sqrt{250} \le m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	2.69/2.66/94.4	4.19/4.14/94.8	4.39/4.35/95.2	3.76/3.74/95.2	3.28/3.27/95.2	2.69/2.67/94.4	4.19/4.13/94.6	4.39/4.32/94.6	3.76/3.68/94.8	3.28/3.21/95.4
$\sqrt{350} \le m_i(0) < \sqrt{500} \to \{\sqrt{350} \le m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	17.2/17.07/94.4	15.37/15.22/95	13.16/13.02/94.8	11.6/11.49/94.8	10.39/10.31/94.6	17.2/17.12/94.6	15.37/15.23/95.2	13.16/12.99/94	11.6/11.42/94.2	10.39/10.23/94.2
$\sqrt{350} \le m_i(0) < \sqrt{500} \to \{\sqrt{500} \le m_i(t)\} \cap T_i^\star > t$	65.81/65.98/95.2	54.5/54.8/95	47.27/47.58/94.8	42.92/43.22/94.8	39.5/39.78/95	65.81/66.23/94	54.5/55.02/94.4	47.27/47.72/95.4	42.92/43.31/94.4	39.5/39.82/94.4
$\sqrt{500} \le m_i(0) \to \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	<0.01/<0.01/93.6	<0.01/<0.01/92.6	<0.01/<0.01/94.8	<0.01/<0.01/94.2	<0.01/<0.01/95.2	<0.01/<0.01/93.6	<0.01/<0.01/93.4	<0.01/<0.01/93.8	<0.01/<0.01/93.6	<0.01/<0.01/94.2
$\sqrt{500} \le m_i(0) \to \{\sqrt{50} \le m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	<0.01/<0.01/94	<0.01/<0.01/93	0.02/0.02/94.6	0.01/0.01/94.6	0.01/0.01/94.4	<0.01/<0.01/93.8	<0.01/<0.01/93.6	0.02/0.02/94.4	0.01/0.01/94.2	0.01/0.01/92.8
$\sqrt{500} \le m_i(0) \to \{\sqrt{100} \le m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	<0.01/<0.01/94	0.11/0.11/93.6	0.3/0.31/95.2	0.24/0.25/95.4	0.2/0.2/94.6	<0.01/<0.01/93.8	0.11/0.11/94	0.3/0.3/95.2	0.24/0.24/95.2	0.2/0.19/94.8
$\sqrt{500} \le m_i(0) \to \{\sqrt{200} \le m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	0.02/0.02/94	0.26/0.26/94	0.53/0.53/95.2	0.45/0.45/95	0.38/0.38/95.2	0.02/0.02/94.4	0.26/0.26/94.4	0.53/0.52/95	0.45/0.43/95	0.38/0.37/94.8
$\sqrt{500} \le m_i(0) \to \{\sqrt{250} \le m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	0.4/0.4/94.8	1.83/1.8/95	2.7/2.67/95	2.33/2.3/95.2	2.04/2.02/94.8	0.4/0.4/94.4	1.83/1.8/94.8	2.7/2.64/94	2.33/2.26/95	2.04/1.98/95
$\sqrt{500} \le m_i(0) \to \{\sqrt{350} \le m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	5.29/5.23/93.6	9.33/9.15/95.2	9.92/9.73/94.6	8.77/8.61/95.2	7.88/7.75/95.8	5.29/5.24/93.8	9.33/9.16/95	9.92/9.7/94.4	8.77/8.55/95	7.88/7.68/95.2
$\sqrt{500} \le m_i(0) \to \{\sqrt{500} \le m_i(t)\} \cap T_i^\star > t$	81.92/81.98/93.6	66.09/66.38/95	55.83/56.17/95	50.88/51.23/94.6	46.97/47.32/94.8	81.92/82.37/92.4	66.09/66.71/93.8	55.83/56.37/94.4	50.88/51.31/93.8	46.97/47.31/95

† Results from 500 replications with each dataset including 1500 individuals. The true marker evolution was based on linear splines with knots at 1 and 5 years since baseline and it was correctly specified in the fitted SREM-CIF-1 and SREM-CIF-2 models. The true parameter values, the mean of posterior medians over the 500 replications and the empirical coverage probability of posterior credible intervals are presented, respectively.

Table 5.9: Simulation study results for transition marker state probabilities (%) by baseline marker state for group 1 ($w = 1$) when the data have been generated by SREM-CIF-2 and there is no misclassification†.

transition	$t=2$	$t=4$	$t=6$	$t=8$	$t=10$	$t=2$	$t=4$	$t=6$	$t=8$	$t=10$
	Results from SREM-CIF-1					Results from SREM-CIF-2				
$m_i(0) < \sqrt{50} \rightarrow \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	1.43/1.45/94.6	0.23/0.22/93.8	0.08/0.08/94.6	0.03/0.02/93.4	<0.01/0.01/94	1.43/1.45/95.4	0.23/0.24/94.4	0.08/0.09/94.4	0.03/0.03/94.4	<0.01/0.01/94.8
$m_i(0) < \sqrt{50} \rightarrow \{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	5.95/6.05/93.8	1.26/1.29/94	0.53/0.53/94.8	0.3/0.29/95.2	0.15/0.14/94.2	5.95/5.97/94.2	1.26/1.31/93.8	0.53/0.57/93.6	0.3/0.34/94	0.15/0.18/95.4
$m_i(0) < \sqrt{50} \rightarrow \{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	24.39/24.68/94	8.11/8.26/92.6	3.94/4/93.2	2.9/2.9/94.6	2.14/2.11/95.6	24.39/24.38/93.6	8.11/8.24/93.2	3.94/4.07/92.8	2.9/3.01/93.6	2.14/2.25/95.2
$m_i(0) < \sqrt{50} \rightarrow \{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	12.33/12.43/93.8	6.5/6.58/93.8	3.61/3.64/92.4	2.88/2.89/94.4	2.34/2.34/95	12.33/12.3/94	6.5/6.53/93.4	3.61/3.65/93.4	2.88/2.93/93.6	2.34/2.39/94.4
$m_i(0) < \sqrt{50} \rightarrow \{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	15.39/15.5/93.4	13.94/14.03/93.8	9.18/9.21/94.8	7.66/7.67/95	6.52/6.51/95.4	15.39/15.36/93.6	13.94/13.9/94.2	9.18/9.19/94.8	7.66/7.69/94.8	6.52/6.56/95.8
$m_i(0) < \sqrt{50} \rightarrow \{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	6.17/6.23/95	14.49/14.5/93.2	13.05/12.99/95.4	11.33/11.27/95	10.01/9.96/94.8	6.17/6.18/94.2	14.49/14.37/92.8	13.05/12.92/94.4	11.33/11.24/94.6	10.01/9.94/94.4
$m_i(0) < \sqrt{50} \rightarrow \{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	0.65/0.67/95.2	7.38/7.4/93	12.55/12.54/93.4	11.27/11.26/93.8	10.27/10.26/94	0.65/0.66/94.8	7.38/7.35/93	12.55/12.47/93.6	11.27/11.2/94.2	10.27/10.22/93.4
$\sqrt{50} \leq m_i(0) < \sqrt{100} \rightarrow \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	0.11/0.11/94	0.07/0.07/93.8	0.05/0.06/94.8	0.02/0.02/95.2	0.01/0.01/94.6	0.11/0.11/94.2	0.07/0.07/94	0.05/0.06/95.8	0.02/0.03/95.2	0.01/0.01/95.8
$\sqrt{50} \leq m_i(0) < \sqrt{100} \rightarrow \{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	1.26/1.28/94.2	0.54/0.57/93.2	0.36/0.38/93.2	0.24/0.25/94.8	0.15/0.16/95.6	1.26/1.26/94.6	0.54/0.56/93	0.36/0.38/94.4	0.24/0.25/95	0.15/0.16/95.4
$\sqrt{50} \leq m_i(0) < \sqrt{100} \rightarrow \{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	13.66/13.76/92.4	5.18/5.32/91	3.12/3.25/91.6	2.43/2.53/92	1.92/2/93.4	13.66/13.67/93.8	5.18/5.26/92.2	3.12/3.21/93.2	2.43/2.5/93.8	1.92/1.99/94
$\sqrt{50} \leq m_i(0) < \sqrt{100} \rightarrow \{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	12.56/12.62/94.8	5.36/5.44/91.6	3.25/3.33/92	2.67/2.75/92.4	2.25/2.32/92.6	12.56/12.57/94.8	5.36/5.4/93.2	3.25/3.29/92.6	2.67/2.72/93.8	2.25/2.29/94
$\sqrt{50} \leq m_i(0) < \sqrt{100} \rightarrow \{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	25.07/25.15/95	14.47/14.59/92.2	9.39/9.51/93	8.01/8.13/93.4	6.97/7.09/93.2	25.07/25.07/94.6	14.47/14.5/94	9.39/9.43/93.6	8.01/8.06/93.8	6.97/7.02/94
$\sqrt{50} \leq m_i(0) < \sqrt{100} \rightarrow \{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	18.18/18.22/94.4	20.3/20.32/93.2	15.93/15.95/93	14.03/14.06/94	12.57/12.63/93.4	18.18/18.17/93.6	20.3/20.24/93	15.93/15.87/92.8	14.03/13.99/93.6	12.57/12.55/93.2
$\sqrt{50} \leq m_i(0) < \sqrt{100} \rightarrow \{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	3.75/3.79/95.2	15.29/15.29/93.8	20.19/20.16/94.6	18.3/18.29/95.2	16.82/16.84/95.8	3.75/3.78/95	15.29/15.26/94	20.19/20.13/95	18.3/18.25/95	16.82/16.79/95.4
$\sqrt{100} \leq m_i(0) < \sqrt{200} \rightarrow \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	0.01/0.01/94.8	0.02/0.02/93	0.03/0.03/95	0.01/0.02/95.6	0.01/0.01/95	0.01/0.01/95	0.02/0.02/94	0.03/0.03/95	0.01/0.02/96	0.01/0.01/96
$\sqrt{100} \leq m_i(0) < \sqrt{200} \rightarrow \{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	0.22/0.23/94.8	0.22/0.23/93	0.21/0.23/93	0.15/0.16/94	0.1/0.11/94	0.22/0.22/94.8	0.22/0.23/94.2	0.21/0.22/95.4	0.15/0.16/95.2	0.1/0.11/95.2
$\sqrt{100} \leq m_i(0) < \sqrt{200} \rightarrow \{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	4.86/4.88/94	2.88/2.95/94.4	2.2/2.29/92.8	1.74/1.83/92.2	1.41/1.5/91.2	4.86/4.86/94.4	2.88/2.91/94.4	2.2/2.25/95	1.74/1.79/94.8	1.41/1.46/94.4
$\sqrt{100} \leq m_i(0) < \sqrt{200} \rightarrow \{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	7.03/7.03/95.2	3.64/3.68/93.2	2.56/2.62/92.2	2.13/2.2/91.4	1.82/1.89/91.4	7.03/7.03/95.6	3.64/3.66/93.8	2.56/2.59/94.8	2.13/2.16/95	1.82/1.85/93.8
$\sqrt{100} \leq m_i(0) < \sqrt{200} \rightarrow \{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	22.35/22.34/94.4	12.05/12.11/92.8	8.34/8.43/93.2	7.18/7.29/92.8	6.31/6.44/91.8	22.35/22.35/94.2	12.05/12.07/93.6	8.34/8.37/94.2	7.18/7.21/94.6	6.31/6.35/94.2
$\sqrt{100} \leq m_i(0) < \sqrt{200} \rightarrow \{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	30.61/30.57/93.2	22.4/22.38/93	16.71/16.69/93.2	14.81/14.84/93.8	13.36/13.43/93.6	30.61/30.62/94.8	22.4/22.37/93	16.71/16.65/92.8	14.81/14.77/93.2	13.36/13.34/93
$\sqrt{100} \leq m_i(0) < \sqrt{200} \rightarrow \{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	14.21/14.2/93.8	25.69/25.65/93.2	28.15/28.1/94.4	25.66/25.64/94.6	23.71/23.74/94.6	14.21/14.22/94	25.69/25.7/94	28.15/28.14/94.4	25.66/25.66/95.4	23.71/23.73/94.2
$\sqrt{200} \leq m_i(0) < \sqrt{250} \rightarrow \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	<0.01/<0.01/94.6	0.01/0.01/93.4	0.01/0.02/93	0.01/0.01/94.2	<0.01/0.01/93.4	<0.01/<0.01/94.8	0.01/0.01/93.8	0.01/0.02/94	0.01/0.01/95	<0.01/<0.01/94.6
$\sqrt{200} \leq m_i(0) < \sqrt{250} \rightarrow \{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	0.03/0.03/94.8	0.09/0.09/92.6	0.12/0.14/92.4	0.09/0.1/92.6	0.06/0.07/92.6	0.03/0.03/94.4	0.09/0.09/93.4	0.12/0.13/94	0.09/0.09/94	0.06/0.07/94.2
$\sqrt{200} \leq m_i(0) < \sqrt{250} \rightarrow \{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	1.29/1.31/94.8	1.52/1.56/94.2	1.51/1.58/93	1.21/1.28/92.4	1/1.06/92	1.29/1.31/94.8	1.52/1.54/93.8	1.51/1.55/94.8	1.21/1.24/93.8	1/1.03/93
$\sqrt{200} \leq m_i(0) < \sqrt{250} \rightarrow \{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	2.88/2.87/94.8	2.31/2.33/94.2	1.95/2/93.4	1.64/1.69/93	1.41/1.46/91.4	2.88/2.88/94.6	2.31/2.31/94.4	1.95/1.97/94.4	1.64/1.66/93.2	1.41/1.43/93.6
$\sqrt{200} \leq m_i(0) < \sqrt{250} \rightarrow \{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	14.34/14.27/95	9.19/9.19/94.6	7.07/7.12/94.2	6.11/6.2/93.6	5.4/5.51/93	14.34/14.31/93.6	9.19/9.18/94	7.07/7.07/94.2	6.11/6.12/94.2	5.4/5.42/94
$\sqrt{200} \leq m_i(0) < \sqrt{250} \rightarrow \{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	33.97/33.83/95	21.77/21.67/93.2	16.32/16.26/93.2	14.52/14.52/93.8	13.15/13.2/94	33.97/33.96/95	21.77/21.71/93	16.32/16.24/93.2	14.52/14.46/93.6	13.15/13.11/93.2
$\sqrt{200} \leq m_i(0) < \sqrt{250} \rightarrow \{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	29.69/29.59/93.6	35.72/35.62/94.2	35.13/35.04/94.4	32.12/32.1/93.8	29.76/29.82/92	29.69/29.7/93.2	35.72/35.77/94.2	35.13/35.17/94.4	32.12/32.17/93.4	29.76/29.84/92.6
$\sqrt{250} \leq m_i(0) < \sqrt{350} \rightarrow \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	<0.01/<0.01/94.2	<0.01/<0.01/93.6	0.01/0.01/93.6	0.01/0.01/93.8	<0.01/<0.01/93.2	<0.01/<0.01/94.6	<0.01/<0.01/94.2	0.01/0.01/94.4	0.01/0.01/94.6	<0.01/<0.01/94.4
$\sqrt{250} \leq m_i(0) < \sqrt{350} \rightarrow \{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	0.01/0.01/94.4	0.04/0.04/94	0.08/0.09/93	0.06/0.06/92.6	0.04/0.05/92.4	0.01/0.01/95	0.04/0.04/94.4	0.08/0.09/94.2	0.06/0.06/94.2	0.04/0.05/94.4
$\sqrt{250} \leq m_i(0) < \sqrt{350} \rightarrow \{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	0.42/0.43/95.2	0.89/0.91/93.6	1.09/1.14/93.8	0.88/0.93/92.8	0.73/0.78/92.6	0.42/0.43/95.2	0.89/0.9/94.6	1.09/1.12/94	0.88/0.9/93.8	0.73/0.75/92.8
$\sqrt{250} \leq m_i(0) < \sqrt{350} \rightarrow \{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	1.2/1.2/94.6	1.52/1.53/94.2	1.52/1.55/94.2	1.28/1.32/93	1.11/1.15/91.6	1.2/1.2/95.2	1.52/1.52/94.2	1.52/1.53/94	1.28/1.29/93.2	1.11/1.12/93
$\sqrt{250} \leq m_i(0) < \sqrt{350} \rightarrow \{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	8.1/8.06/95	6.91/6.9/94.8	5.95/5.98/94.2	5.17/5.22/93.4	4.58/4.66/92.8	8.1/8.09/94.6	6.91/6.89/94.8	5.95/5.94/93.8	5.17/5.16/93.8	4.58/4.58/93.2
$\sqrt{250} \leq m_i(0) < \sqrt{350} \rightarrow \{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	28.95/28.74/94	19.65/19.49/94.8	15.33/15.22/94.6	13.67/13.63/94	12.4/12.42/94.8	28.95/28.88/93.6	19.65/19.54/95.6	15.33/15.21/94.6	13.67/13.57/94.4	12.4/12.33/94
$\sqrt{250} \leq m_i(0) < \sqrt{350} \rightarrow \{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	45.21/44.98/93.2	43.83/43.65/94.2	40.55/40.42/94	37.15/37.11/94	34.5/34.55/93.8	45.21/45.2/94	43.83/43.89/93.4	40.55/40.6/94.2	37.15/37.21/94.6	34.5/34.58/93.8
$\sqrt{350} \leq m_i(0) < \sqrt{500} \rightarrow \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	<0.01/<0.01/94.6	<0.01/<0.01/94.6	<0.01/<0.01/93.4	<0.01/<0.01/93	<0.01/<0.01/93	<0.01/<0.01/94.4	<0.01/<0.01/94.8	<0.01/<0.01/94.6	<0.01/<0.01/94.2	<0.01/<0.01/94
$\sqrt{350} \leq m_i(0) < \sqrt{500} \rightarrow \{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	<0.01/<0.01/95	0.01/0.01/94.6	0.04/0.05/94	0.03/0.03/93.6	0.02/0.03/93.2	<0.01/<0.01/94.6	0.01/0.01/95.2	0.04/0.04/94.2	0.03/0.03/94	0.02/0.02/94
$\sqrt{350} \leq m_i(0) < \sqrt{500} \rightarrow \{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	0.06/0.07/94.8	0.37/0.39/94.2	0.65/0.68/94.8	0.53/0.56/94.2	0.44/0.48/92.8	0.06/0.07/95	0.37/0.38/94.4	0.65/0.67/94.6	0.53/0.54/93.4	0.44/0.46/93.4
$\sqrt{350} \leq m_i(0) < \sqrt{500} \rightarrow \{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	0.25/0.26/94.4	0.76/0.76/94	1.01/1.03/94.8	0.85/0.88/93.6	0.74/0.77/92.4	0.25/0.26/94.8	0.76/0.76/94	1.01/1.02/94.6	0.85/0.86/93.8	0.74/0.75/93.4
$\sqrt{350} \leq m_i(0) < \sqrt{500} \rightarrow \{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	2.67/2.67/94.8	4.18/4.15/94.2	4.42/4.43/94.2	3.85/3.89/94.2	3.42/3.48/93.6	2.67/2.68/94.4	4.18/4.16/94	4.42/4.4/94	3.85/3.83/94.2	3.42/3.42/94
$\sqrt{350} \leq m_i(0) < \sqrt{500} \rightarrow \{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	17.12/16.97/94.2	15.38/15.18/94.6	13.31/13.15/95.4	11.89/11.81/94.6	10.81/10.79/94.4	17.12/17.06/94	15.38/15.23/95	13.31/13.15/95.6	11.89/11.75/94.4	10.81/10.7/92.4
$\sqrt{350} \leq m_i(0) < \sqrt{500} \rightarrow \{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	65.65/65.15/92.6	54.69/54.36/94	47.91/47.69/93.4	44/43.9/93.4	40.96/40.97/93.6	65.65/65.55/93.8	54.69/54.72/94.4	47.91/47.94/93.8	44/44.05/95	40.96/41.03/94
$\sqrt{500} \leq m_i(0) \rightarrow \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	<0.01/<0.01/94.6	<0.01/<0.01/94.6	<0.01/<0.01/93.6	<0.01/<0.01/93.6	<0.01/<0.01/93.4	<0.01/<0.01/94.4	<0.01/<0.01/95	<0.01/<0.01/94.6	<0.01/<0.01/93.4	<0.01/<0.01/94
$\sqrt{500} \leq m_i(0) \rightarrow \{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	<0.01/<0.01/95	<0.01/<0.01/94.8	0.02/0.02/94.4	0.01/0.01/93.4	0.01/0.01/92.8	<0.01/<0.01/95	<0.01/<0.01/94.6	0.02/0.02/94.4	0.01/0.01/94	0.01/0.01/93.4
$\sqrt{500} \leq m_i(0) \rightarrow \{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	<0.01/<0.01/93.8	0.1/0.11/94	0.31/0.32/94.6	0.25/0.27/93.6	0.21/0.23/92.6	<0.01/<0.01/94	0.1/0.11/94	0.31/0.32/94.2	0.25/0.26/94.4	0.21/0.22/94
$\sqrt{500} \leq m_i(0) \rightarrow \{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	0.02/0.02/94.2	0.26/0.26/93.8	0.54/0.55/94.6	0.46/0.47/94.2	0.4/0.42/93.2	0.02/0.02/94	0.26/0.26/94.2	0.54/0.54/94.4	0.46/0.46/94.2	0.4/0.4/94
$\sqrt{500} \leq m_i(0) \rightarrow \{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	0.39/0.41/94.8	1.83/1.82/94.2	2.73/2.73/93.8	2.38/2.4/94.2	2.12/2.16/93	0.39/0.41/94.6	1.83/1.82/94	2.73/2.71/94	2.38/2.37/94	2.12/2.11/94
$\sqrt{500} \leq m_i(0) \rightarrow \{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	5.27/5.26/94.2	9.35/9.18/94.2	10.04/9.87/94.4	8.99/8.89/93.8	8.19/8.14/93.8	5.27/5.29/94.8	9.35/9.21/94.6	10.04/9.87/94.8	8.99/8.84/93.6	8.19/8.06/93.6
$\sqrt{500} \leq m_i(0) \rightarrow \{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	81.83/81.03/92.4	66.4/65.8/93.4	56.6/56.2/93.8	52.13/51.91/93.6	48.63/48.58/94	81.83/81.57/93.8	66.4/66.29/94	56.6/56.54/94.6	52.13/52.1/94.2	48.63/48.64/94

† Results from 500 replications with each dataset including 1500 individuals. The true marker evolution was based on linear splines with knots at 1 and 5 years since baseline and it was correctly specified in the fitted SREM-CIF-1 and SREM-CIF-2 models. The true parameter values, the mean of posterior medians over the 500 replications and the empirical coverage probability of posterior credible intervals are presented, respectively.

Table 5.10: Simulation study results for transition marker state probabilities (%) by baseline marker state for group 0 ($w = 0$) when the data have been generated by SREM-CIF-2 and there is no misclassification†.

transition	$t=2$	$t=4$	$t=6$	$t=8$	$t=10$	$t=2$	$t=4$	$t=6$	$t=8$	$t=10$
	Results from SREM-CIF-1					Results from SREM-CIF-2				
$m_i(0) < \sqrt{50} \to \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	1.47/1.5/94.6	0.23/0.24/94.8	0.08/0.08/95	0.03/0.03/94.8	<0.01/0.01/94.2	1.47/1.48/96	0.23/0.25/95	0.08/0.09/95	0.03/0.03/95	<0.01/0.01/94.4
$m_i(0) < \sqrt{50} \to \{\sqrt{50} \le m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	6.04/6.16/94.4	1.28/1.32/93.6	0.54/0.55/94.2	0.31/0.3/93.8	0.15/0.15/93.8	6.04/6.04/94.4	1.28/1.32/95.4	0.54/0.57/93.2	0.31/0.33/94.6	0.15/0.18/94.4
$m_i(0) < \sqrt{50} \to \{\sqrt{100} \le m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	24.53/24.85/93.4	8.15/8.34/94.2	3.94/4.03/93.2	2.89/2.9/93.6	2.12/2.09/94	24.53/24.46/96	8.15/8.23/94.4	3.94/4.04/93.4	2.89/2.96/93	2.12/2.2/94.4
$m_i(0) < \sqrt{50} \to \{\sqrt{200} \le m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	12.34/12.45/95.8	6.5/6.59/93.4	3.59/3.63/93.2	2.85/2.86/93.4	2.3/2.28/93.6	12.34/12.28/95.4	6.5/6.49/95.2	3.59/3.61/93.8	2.85/2.87/93	2.3/2.32/92.6
$m_i(0) < \sqrt{50} \to \{\sqrt{250} \le m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	15.36/15.48/93.6	13.87/13.96/94.8	9.09/9.11/93.6	7.54/7.52/94.8	6.38/6.32/94.8	15.36/15.29/93	13.87/13.77/95.4	9.09/9.04/94.6	7.54/7.51/94.4	6.38/6.36/94.4
$m_i(0) < \sqrt{50} \to \{\sqrt{350} \le m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	6.14/6.21/93.8	14.34/14.35/94.4	12.84/12.77/95.8	11.08/10.97/94.8	9.74/9.61/94.2	6.14/6.14/93.8	14.34/14.18/93	12.84/12.66/93.8	11.08/10.93/93.8	9.74/9.61/94
$m_i(0) < \sqrt{50} \to \{\sqrt{500} \le m_i(t)\} \cap T_i^\star > t$	0.64/0.66/94.8	7.27/7.29/93.6	12.29/12.26/93.8	10.96/10.9/93.8	9.94/9.85/94.4	0.64/0.66/95.2	7.27/7.22/93.2	12.29/12.17/93.2	10.96/10.85/94.2	9.94/9.83/95.2
$\sqrt{50} \le m_i(0) < \sqrt{100} \to \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	0.11/0.11/94	0.07/0.08/93.2	0.05/0.06/95	0.02/0.03/94.8	0.01/0.01/94	0.11/0.11/94	0.07/0.08/94.4	0.05/0.06/96	0.02/0.03/94.8	0.01/0.01/94.6
$\sqrt{50} \le m_i(0) < \sqrt{100} \to \{\sqrt{50} \le m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	1.26/1.29/93.6	0.54/0.57/93.4	0.36/0.38/95	0.24/0.25/94.4	0.15/0.16/94.4	1.26/1.27/94.2	0.54/0.56/94.8	0.36/0.38/95.8	0.24/0.25/94.6	0.15/0.16/94.4
$\sqrt{50} \le m_i(0) < \sqrt{100} \to \{\sqrt{100} \le m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	13.65/13.76/94	5.16/5.3/93.2	3.09/3.22/92.6	2.39/2.49/93.4	1.89/1.95/94	13.65/13.63/95	5.16/5.21/93.8	3.09/3.16/94	2.39/2.45/93.8	1.89/1.94/93.6
$\sqrt{50} \le m_i(0) < \sqrt{100} \to \{\sqrt{200} \le m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	12.51/12.58/94.8	5.31/5.4/93.8	3.2/3.27/92.8	2.62/2.68/93.2	2.19/2.24/92.4	12.51/12.5/95.6	5.31/5.33/94.2	3.2/3.23/93.4	2.62/2.64/93.4	2.19/2.22/93.4
$\sqrt{50} \le m_i(0) < \sqrt{100} \to \{\sqrt{250} \le m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	24.9/25.01/95.8	14.29/14.4/93.6	9.22/9.31/94	7.82/7.89/94.4	6.77/6.82/94.2	24.9/24.87/95.2	14.29/14.27/94.4	9.22/9.22/95.2	7.82/7.82/93.8	6.77/6.78/94.4
$\sqrt{50} \le m_i(0) < \sqrt{100} \to \{\sqrt{350} \le m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	18.02/18.08/94	19.97/19.99/94.4	15.57/15.55/94	13.63/13.58/94	12.14/12.09/93.8	18.02/17.98/93.6	19.97/19.86/95	15.57/15.45/93.4	13.63/13.52/93.6	12.14/12.06/94.6
$\sqrt{50} \le m_i(0) < \sqrt{100} \to \{\sqrt{500} \le m_i(t)\} \cap T_i^\star > t$	3.71/3.75/95.2	14.99/15/93.8	19.65/19.59/95	17.69/17.6/95.8	16.17/16.08/95.6	3.71/3.73/94.8	14.99/14.93/94.2	19.65/19.53/95	17.69/17.58/95	16.17/16.09/94.8
$\sqrt{100} \le m_i(0) < \sqrt{200} \to \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	0.01/0.01/94.6	0.02/0.03/94	0.03/0.03/95.8	0.01/0.02/94.8	0.01/0.01/94	0.01/0.01/94.8	0.02/0.02/94.2	0.03/0.03/96.8	0.01/0.02/96.2	0.01/0.01/94.6
$\sqrt{100} \le m_i(0) < \sqrt{200} \to \{\sqrt{50} \le m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	0.22/0.23/94.2	0.22/0.23/92.8	0.21/0.23/94.6	0.15/0.16/94.8	0.1/0.11/95	0.22/0.22/94.4	0.22/0.23/93.8	0.21/0.22/96.8	0.15/0.15/96	0.1/0.11/95.6
$\sqrt{100} \le m_i(0) < \sqrt{200} \to \{\sqrt{100} \le m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	4.84/4.86/94	2.85/2.92/94	2.16/2.25/94	1.71/1.79/94.8	1.38/1.45/94.2	4.84/4.83/94	2.85/2.88/95	2.16/2.21/95.2	1.71/1.74/95.8	1.38/1.41/94.6
$\sqrt{100} \le m_i(0) < \sqrt{200} \to \{\sqrt{200} \le m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	6.98/6.99/95	3.59/3.63/93.6	2.51/2.56/93.8	2.08/2.13/94.4	1.77/1.81/93.8	6.98/6.97/95	3.59/3.6/93	2.51/2.53/94.4	2.08/2.1/94.6	1.77/1.78/95
$\sqrt{100} \le m_i(0) < \sqrt{200} \to \{\sqrt{250} \le m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	22.15/22.15/95	11.85/11.9/94.6	8.15/8.21/95.4	6.97/7.04/95.2	6.1/6.16/95.6	22.15/22.12/95	11.85/11.83/94.6	8.15/8.14/95.2	6.97/6.97/95.4	6.1/6.1/96
$\sqrt{100} \le m_i(0) < \sqrt{200} \to \{\sqrt{350} \le m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	30.27/30.26/93.6	21.96/21.94/94.8	16.26/16.21/94.2	14.32/14.28/93.4	12.85/12.81/94.2	30.27/30.23/93.8	21.96/21.88/94.4	16.26/16.15/93.4	14.32/14.23/93.6	12.85/12.77/94.2
$\sqrt{100} \le m_i(0) < \sqrt{200} \to \{\sqrt{500} \le m_i(t)\} \cap T_i^\star > t$	14.03/14.04/94.2	25.11/25.08/96	27.3/27.21/95.6	24.71/24.6/94.2	22.71/22.6/94.2	14.03/14.02/94.2	25.11/25.07/95.8	27.3/27.21/96.6	24.71/24.64/94.6	22.71/22.65/94.8
$\sqrt{200} \le m_i(0) < \sqrt{250} \to \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	<0.01/<0.01/94.8	0.01/0.01/94	0.01/0.02/94.8	0.01/0.01/95.2	<0.01/0.01/95	<0.01/<0.01/95	0.01/0.01/93.6	0.01/0.02/95.4	0.01/0.01/95.6	<0.01/<0.01/96.4
$\sqrt{200} \le m_i(0) < \sqrt{250} \to \{\sqrt{50} \le m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	0.03/0.03/94	0.09/0.09/94.2	0.12/0.14/94.2	0.09/0.1/94.2	0.06/0.07/94.4	0.03/0.03/94.8	0.09/0.09/94.6	0.12/0.13/94.6	0.09/0.09/95.4	0.06/0.07/97
$\sqrt{200} \le m_i(0) < \sqrt{250} \to \{\sqrt{100} \le m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	1.29/1.3/94.8	1.5/1.54/93.6	1.48/1.54/94.6	1.19/1.24/94.4	0.97/1.03/93.6	1.29/1.29/94.8	1.5/1.52/94.6	1.48/1.51/95.6	1.19/1.21/96.8	0.97/0.99/95.6
$\sqrt{200} \le m_i(0) < \sqrt{250} \to \{\sqrt{200} \le m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	2.85/2.85/94.8	2.27/2.29/94.8	1.91/1.94/95.6	1.59/1.63/94	1.36/1.4/94.8	2.85/2.85/94.8	2.27/2.27/95.6	1.91/1.92/96.6	1.59/1.6/96.4	1.36/1.37/96
$\sqrt{200} \le m_i(0) < \sqrt{250} \to \{\sqrt{250} \le m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	14.18/14.13/95	9.01/9.01/94.2	6.88/6.91/95.4	5.92/5.96/95.6	5.21/5.26/96.4	14.18/14.14/94.8	9.01/8.98/94.4	6.88/6.86/95.8	5.92/5.9/95.4	5.21/5.2/95.8
$\sqrt{200} \le m_i(0) < \sqrt{250} \to \{\sqrt{350} \le m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	33.55/33.44/95.2	21.29/21.2/94.4	15.84/15.74/95	14/13.93/94.4	12.61/12.56/94	33.55/33.49/94.8	21.29/21.18/95	15.84/15.71/94.2	14/13.89/94.8	12.61/12.52/94
$\sqrt{200} \le m_i(0) < \sqrt{250} \to \{\sqrt{500} \le m_i(t)\} \cap T_i^\star > t$	29.28/29.22/92.8	34.86/34.78/92.8	34/33.87/92.6	30.87/30.75/92.8	28.45/28.34/92	29.28/29.26/92.8	34.86/34.83/93	34/33.94/93.2	30.87/30.82/93.2	28.45/28.42/92.2
$\sqrt{250} \le m_i(0) < \sqrt{350} \to \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	<0.01/<0.01/94.2	<0.01/<0.01/94	0.01/0.01/93.4	0.01/0.01/94.6	<0.01/<0.01/95.8	<0.01/<0.01/94.4	<0.01/<0.01/93.6	0.01/0.01/95.2	0.01/0.01/95.4	<0.01/<0.01/95.8
$\sqrt{250} \le m_i(0) < \sqrt{350} \to \{\sqrt{50} \le m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	0.01/0.01/94.2	0.04/0.04/94	0.08/0.09/93.2	0.06/0.06/93.6	0.04/0.05/94	0.01/0.01/94.4	0.04/0.04/94.2	0.08/0.08/95	0.06/0.06/95.2	0.04/0.04/96.4
$\sqrt{250} \le m_i(0) < \sqrt{350} \to \{\sqrt{100} \le m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	0.42/0.43/94.4	0.88/0.89/94	1.07/1.11/94.8	0.86/0.9/94.4	0.71/0.75/93.6	0.42/0.42/94.8	0.88/0.89/94.4	1.07/1.09/96	0.86/0.88/96.4	0.71/0.72/95.8
$\sqrt{250} \le m_i(0) < \sqrt{350} \to \{\sqrt{200} \le m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	1.18/1.19/94.4	1.49/1.5/93.6	1.48/1.51/96	1.24/1.27/95.4	1.07/1.1/94.6	1.18/1.19/94.6	1.49/1.49/94.2	1.48/1.49/96	1.24/1.25/96.6	1.07/1.07/96.2
$\sqrt{250} \le m_i(0) < \sqrt{350} \to \{\sqrt{250} \le m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	8.01/7.98/94.4	6.77/6.75/96	5.79/5.8/95.8	4.99/5.02/96.2	4.41/4.44/96.4	8.01/7.99/94.6	6.77/6.73/95.4	5.79/5.76/96.2	4.99/4.97/97.4	4.41/4.39/97.6
$\sqrt{250} \le m_i(0) < \sqrt{350} \to \{\sqrt{350} \le m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	28.57/28.4/94	19.19/19.04/94.4	14.86/14.72/94.4	13.16/13.06/95	11.88/11.81/95.6	28.57/28.46/94.4	19.19/19.05/94.4	14.86/14.7/94.2	13.16/13.02/94.4	11.88/11.76/95.4
$\sqrt{250} \le m_i(0) < \sqrt{350} \to \{\sqrt{500} \le m_i(t)\} \cap T_i^\star > t$	44.56/44.39/93.4	42.73/42.58/93.2	39.2/39.04/93	35.67/35.52/93.4	32.93/32.82/92.2	44.56/44.51/94	42.73/42.7/94.2	39.2/39.14/93.2	35.67/35.62/93.8	32.93/32.9/92.8
$\sqrt{350} \le m_i(0) < \sqrt{500} \to \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	<0.01/<0.01/94.4	<0.01/<0.01/94.4	<0.01/<0.01/93.4	<0.01/<0.01/94	<0.01/<0.01/94.8	<0.01/<0.01/94.2	<0.01/<0.01/94.6	<0.01/<0.01/95	<0.01/<0.01/94.4	<0.01/<0.01/95.8
$\sqrt{350} \le m_i(0) < \sqrt{500} \to \{\sqrt{50} \le m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	<0.01/<0.01/94.8	0.01/0.01/94.2	0.04/0.04/93.4	0.03/0.03/94	0.02/0.03/93.2	<0.01/<0.01/94.6	0.01/0.01/94.4	0.04/0.04/94.8	0.03/0.03/95	0.02/0.02/95.2
$\sqrt{350} \le m_i(0) < \sqrt{500} \to \{\sqrt{100} \le m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	0.06/0.07/94.4	0.37/0.38/94	0.64/0.67/95	0.52/0.55/93.6	0.43/0.46/93.6	0.06/0.07/94.4	0.37/0.37/94	0.64/0.65/95.4	0.52/0.53/95.6	0.43/0.44/95
$\sqrt{350} \le m_i(0) < \sqrt{500} \to \{\sqrt{200} \le m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	0.25/0.25/94.4	0.74/0.75/94.2	0.98/1/94.6	0.83/0.85/94.6	0.71/0.74/93.4	0.25/0.25/94.2	0.74/0.74/93.6	0.98/0.99/94.8	0.83/0.83/95.2	0.71/0.72/94
$\sqrt{350} \le m_i(0) < \sqrt{500} \to \{\sqrt{250} \le m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	2.63/2.64/94.6	4.08/4.06/94.6	4.29/4.29/95.4	3.71/3.73/94.6	3.29/3.31/94.2	2.63/2.65/94.6	4.08/4.05/94.2	4.29/4.26/95.4	3.71/3.69/95.2	3.29/3.27/94.8
$\sqrt{350} \le m_i(0) < \sqrt{500} \to \{\sqrt{350} \le m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	16.88/16.75/94.4	15/14.82/94.6	12.88/12.71/95.4	11.43/11.31/95.2	10.34/10.25/95.6	16.88/16.8/94	15/14.83/94.8	12.88/12.69/95.6	11.43/11.27/95.2	10.34/10.2/95.2
$\sqrt{350} \le m_i(0) < \sqrt{500} \to \{\sqrt{500} \le m_i(t)\} \cap T_i^\star > t$	64.67/64.26/93.2	53.28/53/93.4	46.26/46.03/92.8	42.19/42/93.2	39.05/38.9/92.8	64.67/64.5/93.6	53.28/53.2/93.6	46.26/46.18/93.8	42.19/42.12/93.8	39.05/38.99/92.6
$\sqrt{500} \le m_i(0) \to \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	<0.01/<0.01/94.6	<0.01/<0.01/94.8	<0.01/<0.01/94.4	<0.01/<0.01/93.6	<0.01/<0.01/93.8	<0.01/<0.01/94.2	<0.01/<0.01/95.2	<0.01/<0.01/94.6	<0.01/<0.01/95	<0.01/<0.01/95
$\sqrt{500} \le m_i(0) \to \{\sqrt{50} \le m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	<0.01/<0.01/94.2	<0.01/<0.01/94.8	0.01/0.02/95	0.01/0.01/93.4	0.01/0.01/93	<0.01/<0.01/94.6	<0.01/<0.01/95.4	0.01/0.02/94.8	0.01/0.01/95.2	0.01/0.01/95
$\sqrt{500} \le m_i(0) \to \{\sqrt{100} \le m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	<0.01/<0.01/94.4	0.1/0.11/94.4	0.3/0.31/94.4	0.24/0.26/93.6	0.2/0.22/93.2	<0.01/<0.01/94.2	0.1/0.11/94.8	0.3/0.31/95	0.24/0.25/95.4	0.2/0.21/94.6
$\sqrt{500} \le m_i(0) \to \{\sqrt{200} \le m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	0.02/0.02/94.4	0.25/0.26/94.4	0.52/0.53/94.6	0.44/0.45/93.8	0.38/0.4/93.6	0.02/0.02/94.2	0.25/0.26/94.8	0.52/0.53/94.8	0.44/0.44/93.8	0.38/0.39/94
$\sqrt{500} \le m_i(0) \to \{\sqrt{250} \le m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	0.39/0.4/94.8	1.78/1.78/94.2	2.64/2.64/94.2	2.29/2.3/93.8	2.03/2.05/93.6	0.39/0.4/94.8	1.78/1.77/94.2	2.64/2.62/94.8	2.29/2.27/94.4	2.03/2.02/93.8
$\sqrt{500} \le m_i(0) \to \{\sqrt{350} \le m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	5.19/5.19/94.6	9.11/8.95/95	9.7/9.53/95.2	8.63/8.5/94.6	7.82/7.73/94	5.19/5.21/94.8	9.11/8.96/95	9.7/9.51/95.2	8.63/8.46/94.6	7.82/7.67/94
$\sqrt{500} \le m_i(0) \to \{\sqrt{500} \le m_i(t)\} \cap T_i^\star > t$	80.59/79.9/93.2	64.65/64.12/94	54.62/54.23/93.4	49.95/49.65/93.6	46.34/46.12/93.8	80.59/80.26/94	64.65/64.43/93.6	54.62/54.44/93.6	49.95/49.78/94	46.34/46.19/94.2

† Results from 500 replications with each dataset including 1500 individuals. The true marker evolution was based on linear splines with knots at 1 and 5 years since baseline and it was correctly specified in the fitted SREM-CIF-1 and SREM-CIF-2 models. The true parameter values, the mean of posterior medians over the 500 replications and the empirical coverage probability of posterior credible intervals are presented, respectively.

5.6.1.3 Simulation study results for population-averaged CIFs by baseline marker state

Results for population-averaged CIFs by baseline marker state are provided in Figures 5.1-5.4. In general, both models, even when misspecified, led to nearly unbiased estimates with satisfactory coverage rates. However, there were few cases where the coverage probabilities from the misspecified model were lower than the nominal level. For example, when data were generated from the SREM-CIF-1 model, the coverage probability based on the SREM-CIF-2 model of the population-averaged CIF of cause 1, conditional on the $m_i(0) > \sqrt{500}$ initial state, reduced to 88% (Figure 5.2).

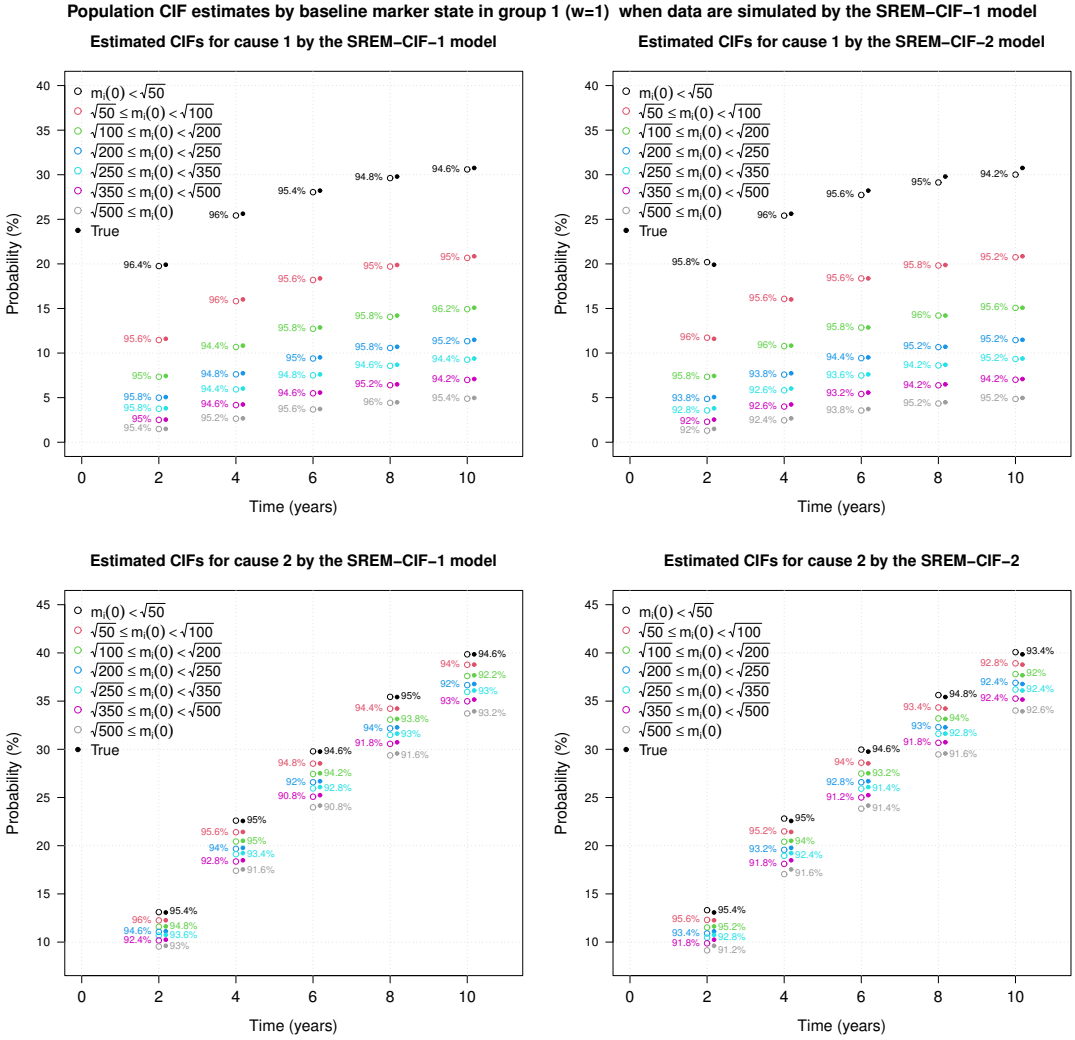


Figure 5.1: Simulation study results for population-averaged CIF estimates by baseline marker state for group 1 when data are simulated under the SREM-CIF-1 model and there is no misclassification. CIF is estimated at certain years since baseline. Open circles show the empirical estimates based on posterior medians over 500 replications whereas closed circles show the true values. Shown are also the corresponding empirical coverage probabilities. The true marker is based on linear splines with knots at 1 and 5 years since baseline and it is correctly specified when fitting SREM-CIF-1 and SREM-CIF-2 models.

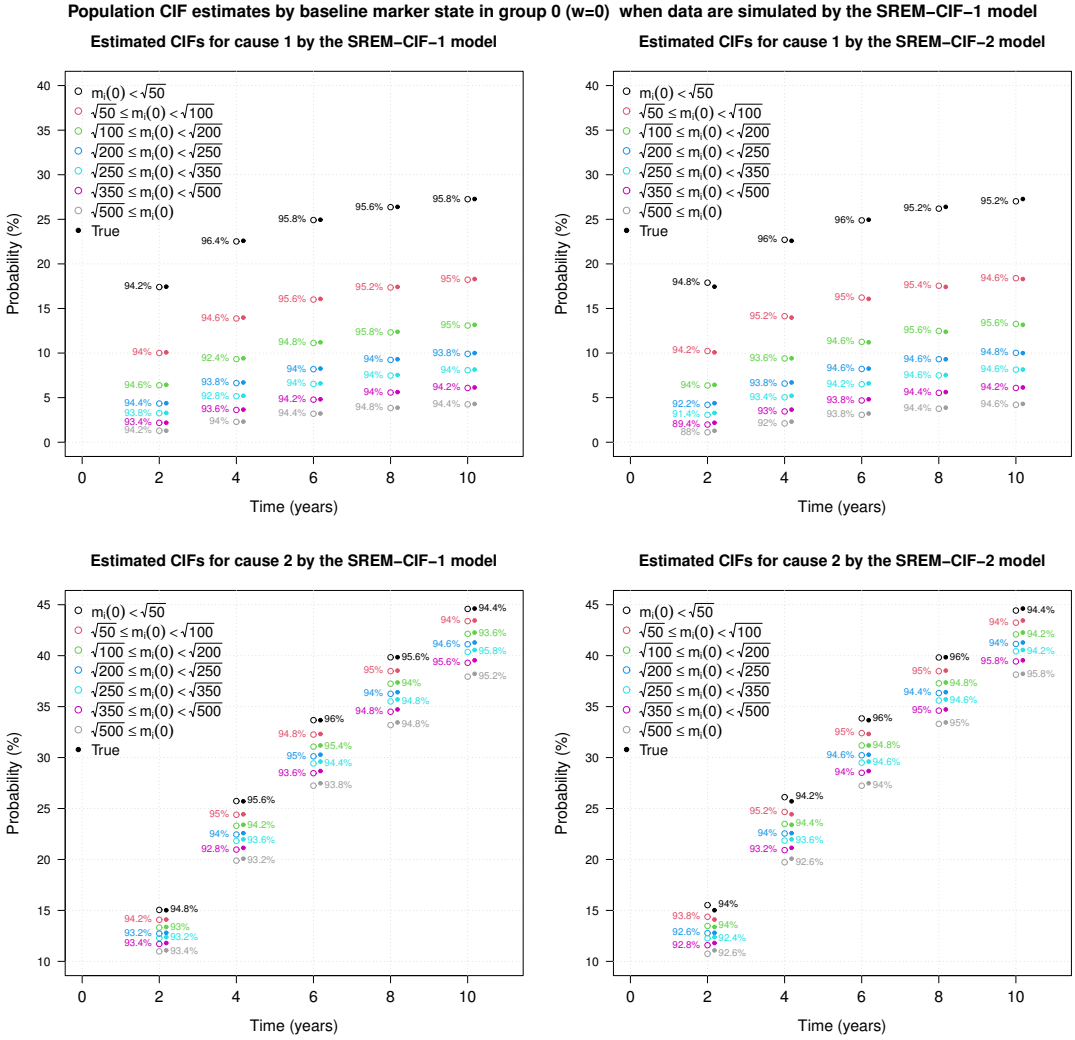


Figure 5.2: Simulation study results for population-averaged CIF estimates by baseline marker state for group 0 when data are simulated under the SREM-CIF-1 model and there is no misclassification. CIF is estimated at certain years since baseline. Open circles show the empirical estimates based on posterior medians over 500 replications whereas closed circles show the true values. Shown are also the corresponding empirical coverage probabilities. The true marker is based on linear splines with knots at 1 and 5 years since baseline and it is correctly specified when fitting SREM-CIF-1 and SREM-CIF-2 models.

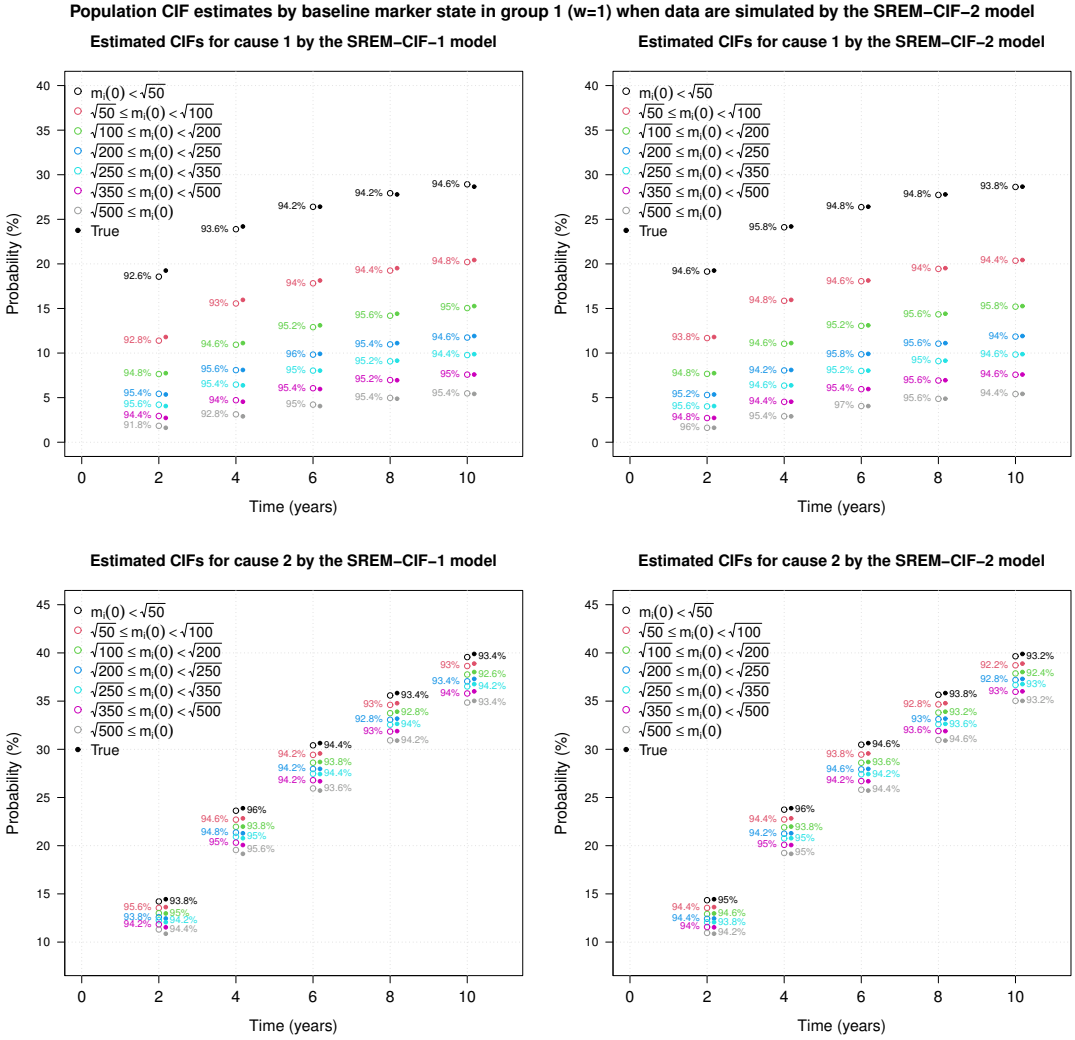


Figure 5.3: Simulation study results for population-averaged CIF estimates by baseline marker state for group 1 when data are simulated under the SREM-CIF-2 model and there is no misclassification. CIF is estimated at certain years since baseline. Open circles show the empirical estimates based on posterior medians over 500 replications whereas closed circles show the true values. Shown are also the corresponding empirical coverage probabilities. The true marker is based on linear splines with knots at 1 and 5 years since baseline and it is correctly specified when fitting SREM-CIF-1 and SREM-CIF-2 models.

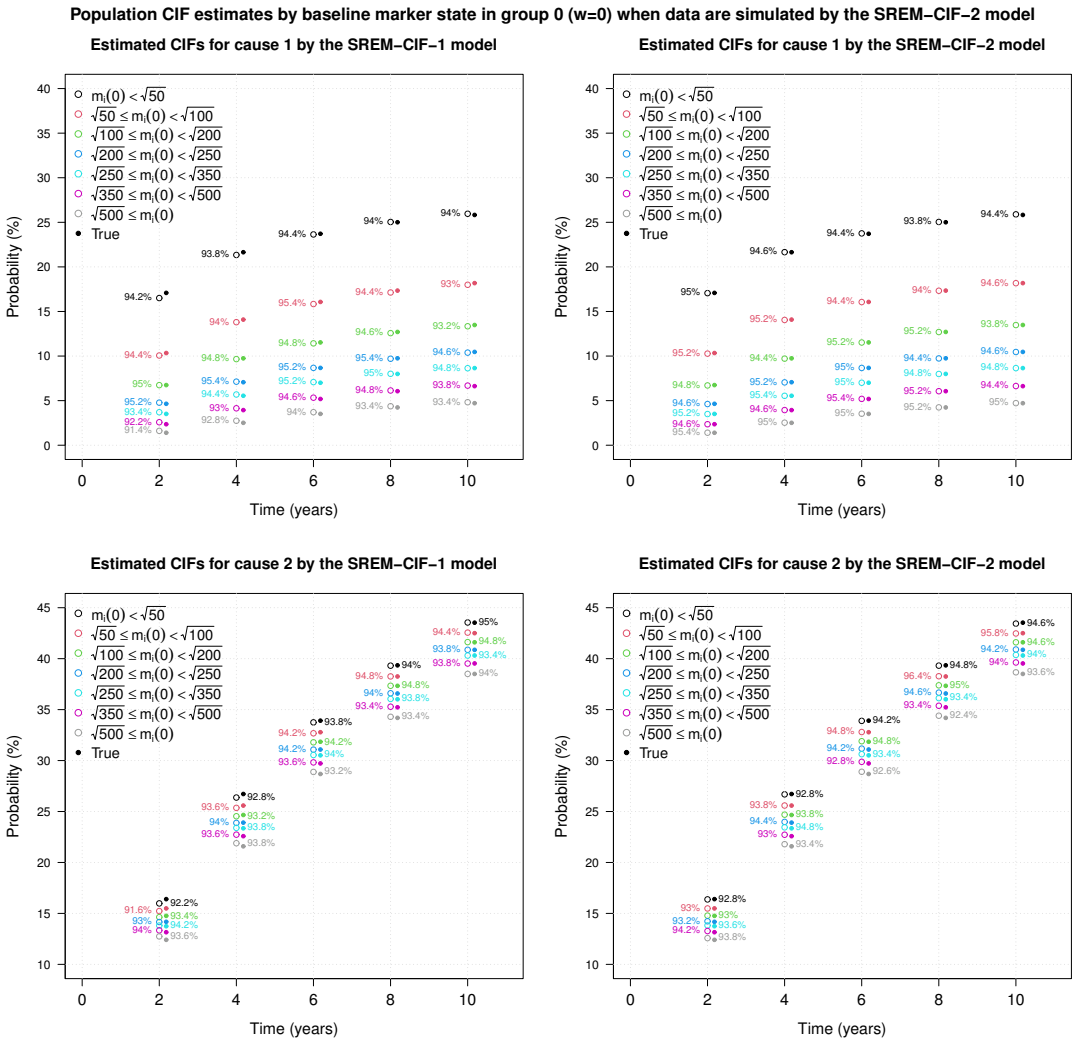


Figure 5.4: Simulation study results for population-averaged CIF estimates by baseline marker state for group 0 when data are simulated under the SREM-CIF-2 model and there is no misclassification. CIF is estimated at certain years since baseline. Open circles show the empirical estimates based on posterior medians over 500 replications whereas closed circles show the true values. Shown are also the corresponding empirical coverage probabilities. The true marker is based on linear splines with knots at 1 and 5 years since baseline and it is correctly specified when fitting SREM-CIF-1 and SREM-CIF-2 models.

5.6.2 Simulation study under misclassification of failure cause

An additional simulation study was performed to assess the model performance when there is failure cause misclassification under the same data generating mechanisms, but including 2000 subjects per dataset. We assumed non-differential misclassification probabilities $\pi_{11} = 0.75$ and $\pi_{22} = 0.90$, that is, the probabilities of observing failure cause 1 and 2 given a true failure cause 1 and 2 were π_{11} and π_{22}, respectively. The basic results are presented in Tables 5.11 and 5.12. Most findings were similar; the models produced estimates with small biases and acceptable coverage rates, with the same conclusion holding for the estimates of the misclassification parameters π_{11} and π_{22}. Both models were able to provide accurate estimates for the population-averaged CIFs and marker states probabilities despite the fact that the survival submodels were misspecified. The performance of the DIC criterion in identifying the true model reduced to some extent to 74.2% and 56.8%, under the SREM-CIF-1 and SREM-CIF-2 scenarios, respectively. Similarly to the no misclassification case, the performance of the DIC criterion substantially improved when including 8000 individuals, becoming 88.0% and 78.0% under the SREM-CIF-1 and SREM-CIF-2 scenarios, respectively.

Table 5.11: Simulation study results from fitted SREM-CIF-1 and SREM-CIF-2 models when the data have been generated by the SREM-CIF-1 model under failure cause misclassification

Parameter	True	Median	Bias	ASD	MCSD	Coverage	Median	Bias	ASD	MCSD	Coverage
Longitudinal	Results from SREM-CIF-1						Results from SREM-CIF-2				
Intercept	12.850	12.856	0.006	0.126	0.122	94.200	12.856	0.006	0.126	0.122	94.000
Slope1 (β_1)	6.030	6.027	-0.003	0.109	0.104	95.600	6.020	-0.010	0.109	0.104	95.200
Slope2 (β_2)	0.770	0.769	-0.001	0.031	0.030	94.800	0.767	-0.003	0.031	0.030	95.000
Slope3 (β_3)	0.000	-0.001	-0.001	0.017	0.017	94.200	-0.001	-0.001	0.017	0.017	94.200
Cause1 (e.g. death)											
"True" marker value (α_1)	-0.160	-0.161	-0.001	0.016	0.017	94.600	-0.182		0.019	0.020	
Binary covariate (γ_1)	0.150	0.147	-0.003	0.147	0.149	94.200	0.159		0.167	0.168	
CIF1 $t = 2$, $w = 1$	8.175	8.122	-0.053	0.983	0.972	94.600	8.102	-0.073	0.998	0.983	94.800
CIF1 $t = 4$, $w = 1$	11.533	11.351	-0.182	1.278	1.286	95.200	11.328	-0.205	1.291	1.290	95.200
CIF1 $t = 6$, $w = 1$	13.495	13.255	-0.240	1.462	1.475	94.400	13.223	-0.272	1.468	1.471	94.200
CIF1 $t = 8$, $w = 1$	14.774	14.488	-0.286	1.584	1.598	94.800	14.456	-0.318	1.585	1.585	94.400
CIF1 $t = 10$, $w = 1$	15.604	15.246	-0.357	1.667	1.693	92.600	15.210	-0.394	1.664	1.677	93.000
CIF1 $t = 2$, $w = 0$	7.106	7.091	-0.015	0.928	0.922	95.200	7.065	-0.041	0.933	0.932	95.000
CIF1 $t = 4$, $w = 0$	10.062	9.943	-0.118	1.224	1.234	95.200	9.941	-0.120	1.225	1.241	94.600
CIF1 $t = 6$, $w = 0$	11.799	11.636	-0.163	1.407	1.418	95.600	11.643	-0.156	1.402	1.422	94.600
CIF1 $t = 8$, $w = 0$	12.935	12.733	-0.202	1.529	1.544	94.600	12.750	-0.186	1.520	1.543	94.000
CIF1 $t = 10$, $w = 0$	13.673	13.410	-0.263	1.613	1.642	94.600	13.433	-0.240	1.600	1.634	93.400
Cause2 (e.g. disengagement)											
"True" marker value (α_2)	-0.020	-0.021	-0.001	0.010	0.010	94.800	-0.026		0.012	0.012	
Binary covariate (γ_2)	-0.150	-0.152	-0.002	0.088	0.087	94.600	-0.184		0.108	0.106	
CIF2 $t = 2$, $w = 1$	11.525	11.500	-0.025	1.002	1.038	92.800	11.456	-0.068	1.034	1.068	93.600
CIF2 $t = 4$, $w = 1$	20.352	20.325	-0.026	1.436	1.507	93.000	20.321	-0.030	1.476	1.527	93.600
CIF2 $t = 6$, $w = 1$	27.328	27.289	-0.039	1.720	1.799	92.000	27.369	0.041	1.744	1.806	92.800
CIF2 $t = 8$, $w = 1$	32.937	32.951	0.014	1.924	1.986	92.600	33.120	0.183	1.919	1.960	91.400
CIF2 $t = 10$, $w = 1$	37.431	37.514	0.083	2.076	2.145	92.400	37.775	0.343	2.041	2.089	92.200
CIF2 $t = 2$, $w = 0$	13.260	13.245	-0.015	1.082	1.109	94.000	13.441	0.181	1.121	1.137	95.000
CIF2 $t = 4$, $w = 0$	23.228	23.224	-0.004	1.503	1.560	94.400	23.434	0.206	1.539	1.577	94.000
CIF2 $t = 6$, $w = 0$	30.982	30.975	-0.007	1.767	1.845	93.800	31.135	0.152	1.778	1.837	94.000
CIF2 $t = 8$, $w = 0$	37.130	37.189	0.059	1.949	2.056	93.400	37.270	0.140	1.926	2.019	93.000
CIF2 $t = 10$, $w = 0$	41.997	42.143	0.146	2.083	2.168	92.800	42.138	0.141	2.029	2.100	93.000
π_{11}	75.000	73.873	-1.127	4.829	4.884	95.000	73.968	-1.032	4.815	4.881	94.800
π_{22}	90.000	89.160	-0.840	1.935	2.051	91.600	89.079	-0.921	1.934	2.046	91.600

† Results from 500 replications with each dataset including 2000 individuals. The true marker evolution was based on linear splines with knots at 1 and 5 years since baseline and it was correctly specified in the fitted SREM-CIF-1 and SREM-CIF-2 models. "True" denotes the true parameter values; "Median" the mean of posterior medians over the 500 replications; "Bias" the mean bias for posterior median estimates; "ASD" the average posterior standard deviation, "MCSD" the empirical Monte carlo deviation of estimates and "Coverage" the empirical coverage probability of posterior credible intervals.

Table 5.12: Simulation study results from fitted SREM-CIF-1 and SREM-CIF-2 models when the data have been generated by the SREM-CIF-2 model under failure cause misclassification

Parameter	True	Median	Bias	ASD	MCSD	Coverage	Median	Bias	ASD	MCSD	Coverage
Longitudinal	Results from SREM-CIF-1						Results from SREM-CIF-2				
Intercept	12.850	12.846	-0.004	0.126	0.126	95.600	12.846	-0.004	0.126	0.126	96.000
Slope1 (β_1)	6.030	6.034	0.004	0.110	0.108	95.400	6.028	-0.002	0.110	0.108	96.000
Slope2 (β_2)	0.770	0.772	0.002	0.031	0.032	93.400	0.770	-0.000	0.031	0.032	93.400
Slope3 (β_3)	0.000	0.001	0.001	0.017	0.017	95.400	0.001	0.001	0.017	0.017	95.600
Cause1 (e.g. death)											
"True" marker value (α_1)	-0.160	-0.143		0.016	0.016		-0.163	-0.003	0.019	0.019	95.600
Binary covariate (γ_1)	0.150	0.141		0.150	0.164		0.159	0.009	0.168	0.179	93.600
CIF1 $t = 2$, $w = 1$	8.319	8.263	-0.056	1.018	1.093	92.600	8.296	-0.023	1.039	1.100	93.000
CIF1 $t = 4$, $w = 1$	11.572	11.393	-0.179	1.315	1.387	91.400	11.444	-0.128	1.333	1.379	93.000
CIF1 $t = 6$, $w = 1$	13.469	13.290	-0.179	1.500	1.596	92.000	13.339	-0.130	1.512	1.579	92.600
CIF1 $t = 8$, $w = 1$	14.709	14.510	-0.200	1.623	1.760	91.800	14.567	-0.143	1.632	1.733	92.400
CIF1 $t = 10$, $w = 1$	15.521	15.264	-0.256	1.707	1.881	90.800	15.315	-0.206	1.713	1.839	92.200
CIF1 $t = 2$, $w = 0$	7.288	7.239	-0.049	0.960	0.992	94.400	7.219	-0.070	0.969	0.989	94.200
CIF1 $t = 4$, $w = 0$	10.196	10.015	-0.181	1.255	1.288	93.600	10.016	-0.180	1.262	1.276	94.400
CIF1 $t = 6$, $w = 0$	11.904	11.699	-0.205	1.438	1.472	93.400	11.711	-0.193	1.440	1.451	94.800
CIF1 $t = 8$, $w = 0$	13.027	12.790	-0.237	1.560	1.588	93.800	12.815	-0.213	1.558	1.558	94.400
CIF1 $t = 10$, $w = 0$	13.765	13.461	-0.304	1.643	1.677	94.000	13.492	-0.273	1.638	1.641	94.600
Cause2 (e.g. disengagement)											
"True" marker value (α_2)	-0.020	-0.016		0.010	0.010		-0.019	0.001	0.012	0.012	95.400
Binary covariate (γ_2)	-0.150	-0.121		0.088	0.091		-0.152	-0.002	0.108	0.109	93.600
CIF2 $t = 2$, $w = 1$	12.861	12.924	0.063	1.074	1.129	92.600	12.846	-0.015	1.111	1.154	91.600
CIF2 $t = 4$, $w = 1$	21.826	21.935	0.110	1.504	1.573	93.000	21.873	0.047	1.540	1.600	93.000
CIF2 $t = 6$, $w = 1$	28.518	28.568	0.050	1.767	1.834	95.200	28.557	0.039	1.787	1.851	94.800
CIF2 $t = 8$, $w = 1$	33.726	33.789	0.062	1.951	2.083	93.400	33.849	0.122	1.949	2.065	92.800
CIF2 $t = 10$, $w = 1$	37.837	37.932	0.095	2.090	2.258	91.400	38.046	0.209	2.064	2.201	91.600
CIF2 $t = 2$, $w = 0$	14.636	14.462	-0.174	1.140	1.202	93.600	14.627	-0.009	1.184	1.251	93.200
CIF2 $t = 4$, $w = 0$	24.488	24.385	-0.103	1.555	1.604	94.400	24.551	0.062	1.592	1.641	95.000
CIF2 $t = 6$, $w = 0$	31.666	31.590	-0.076	1.799	1.870	94.400	31.723	0.057	1.814	1.871	94.600
CIF2 $t = 8$, $w = 0$	37.150	37.203	0.053	1.963	2.044	93.200	37.288	0.139	1.950	2.009	93.400
CIF2 $t = 10$, $w = 0$	41.417	41.613	0.196	2.088	2.192	93.400	41.649	0.232	2.043	2.124	93.400
π_{11}	75.000	73.856	-1.144	4.890	5.123	91.800	73.819	-1.181	4.881	5.104	91.800
π_{22}	90.000	89.096	-0.904	1.982	1.969	92.600	89.032	-0.968	1.977	1.963	92.600

† Results from 500 replications with each dataset including 2000 individuals. The true marker evolution was based on linear splines with knots at 1 and 5 years since baseline and it was correctly specified in the fitted SREM-CIF-1 and SREM-CIF-2 models. "True" denotes the true parameter values; "Median" the mean of posterior medians over the 500 replications; "Bias" the mean bias for posterior median estimates; "ASD" the average posterior standard deviation, "MCSD" the empirical Monte carlo deviation of estimates and "Coverage" the empirical coverage probability of posterior credible intervals.

5.6.2.1 Simulation study results for marker state probabilities under misclassified failure cause

In this subsection, we present results for marker state probabilities in Tables 5.13-5.16. Both models, even when misspecified, yielded estimates with small biases and adequate coverage rates (91.0-96.6%).

Table 5.13: Simulation study results for marker state probabilities (%) for group 1 ($w = 1$) when the data have been generated by SREM-CIF-1 and there is misclassification‡.

State	Est	$t=0$	$t=2$	$t=4$	$t=6$	$t=8$	$t=10$	$t=0$	$t=2$	$t=4$	$t=6$	$t=8$	$t=10$
		Results from SREM-CIF-1						Results from SREM-CIF-2					
$\{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	True	12.431	0.183	0.037	0.020	0.005	0.001	12.431	0.183	0.037	0.020	0.005	0.001
$\{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	Median	12.425	0.184	0.040	0.023	0.007	0.002	12.422	0.186	0.043	0.025	0.008	0.002
$\{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	Bias	-0.006	0.000	0.003	0.003	0.002	0.001	-0.009	0.003	0.006	0.005	0.003	0.001
$\{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	Coverage	94.600	94.400	92.400	94.200	93.200	93.400	95.000	95.400	91.600	91.400	91.000	93.400
$\{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	True	16.038	1.007	0.307	0.190	0.103	0.050	16.038	1.007	0.307	0.190	0.103	0.050
$\{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	Median	16.008	1.009	0.319	0.204	0.114	0.060	16.010	0.997	0.322	0.207	0.118	0.063
$\{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	Bias	-0.030	0.001	0.012	0.014	0.011	0.009	-0.028	-0.010	0.015	0.017	0.016	0.013
$\{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	Coverage	94.800	94.000	92.400	94.000	92.800	92.200	94.400	95.000	92.200	92.800	91.200	92.200
$\{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	True	31.710	7.103	3.098	2.040	1.511	1.127	31.710	7.103	3.098	2.040	1.511	1.127
$\{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	Median	31.675	7.102	3.148	2.103	1.567	1.183	31.684	7.044	3.124	2.087	1.556	1.178
$\{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	Bias	-0.035	-0.001	0.049	0.063	0.056	0.056	-0.026	-0.059	0.026	0.047	0.045	0.051
$\{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	Coverage	93.400	94.200	91.600	94.000	93.200	92.800	94.200	94.000	93.000	94.200	93.200	93.600
$\{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	True	12.103	6.468	3.468	2.364	1.909	1.569	12.103	6.468	3.468	2.364	1.909	1.569
$\{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	Median	12.097	6.463	3.490	2.398	1.943	1.605	12.098	6.441	3.467	2.377	1.920	1.585
$\{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	Bias	-0.006	-0.005	0.022	0.034	0.034	0.036	-0.005	-0.028	-0.001	0.013	0.011	0.016
$\{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	Coverage	95.200	95.400	94.200	94.200	92.200	92.200	95.200	94.000	94.000	94.200	92.600	94.200
$\{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	True	15.610	16.626	10.745	7.677	6.502	5.607	15.610	16.626	10.745	7.677	6.502	5.607
$\{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	Median	15.612	16.609	10.757	7.712	6.545	5.657	15.610	16.605	10.716	7.664	6.484	5.594
$\{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	Bias	0.002	-0.017	0.012	0.035	0.043	0.051	0.001	-0.021	-0.029	-0.013	-0.018	-0.013
$\{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	Coverage	94.600	93.400	95.200	93.800	94.200	94.800	93.800	93.200	94.800	94.200	94.000	95.000
$\{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	True	9.229	24.230	20.064	15.723	13.847	12.389	9.229	24.230	20.064	15.723	13.847	12.389
$\{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	Median	9.249	24.209	20.019	15.684	13.822	12.373	9.244	24.263	20.016	15.647	13.757	12.289
$\{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	Bias	0.020	-0.021	-0.046	-0.039	-0.026	-0.015	0.015	0.033	-0.048	-0.075	-0.090	-0.100
$\{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	Coverage	93.800	94.800	94.400	94.200	94.400	94.200	93.600	94.400	94.400	93.600	94.000	93.400
$\{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	True	2.880	24.682	30.397	31.163	28.411	26.223	2.880	24.682	30.397	31.163	28.411	26.223
$\{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	Median	2.907	24.728	30.478	31.256	28.492	26.283	2.902	24.826	30.583	31.323	28.499	26.229
$\{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	Bias	0.027	0.046	0.082	0.093	0.080	0.060	0.022	0.144	0.187	0.159	0.088	0.006
$\{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	Coverage	92.800	93.800	93.800	94.000	94.000	94.000	93.400	92.800	93.200	92.600	93.000	94.000

† Results from 500 replications with each dataset including 2000 individuals. The true marker evolution was based on linear splines with knots at 1 and 5 years since baseline and it was correctly specified in the fitted SREM-CIF-1 and SREM-CIF-2 models. "True" denotes the true parameter values; "Median" the mean of posterior medians over the 500 replications; "Bias" the mean bias for posterior median estimates; and "Coverage" the empirical coverage probability of posterior credible intervals.

Table 5.14: Simulation study results for marker state probabilities (%) for group 0 ($w = 0$) when the data have been generated by SREM-CIF-1 and there is misclassification†.

State	Est	$t=0$	$t=2$	$t=4$	$t=6$	$t=8$	$t=10$	$t=0$	$t=2$	$t=4$	$t=6$	$t=8$	$t=10$
		Results from SREM-CIF-1						Results from SREM-CIF-2					
$\{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	True	12.431	0.193	0.039	0.021	0.005	0.001	12.431	0.193	0.039	0.021	0.005	0.001
$\{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	Median	12.425	0.193	0.042	0.024	0.007	0.002	12.422	0.190	0.043	0.024	0.007	0.002
$\{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	Bias	-0.006	-0.001	0.003	0.003	0.002	0.001	-0.009	-0.003	0.004	0.004	0.002	0.001
$\{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	Coverage	94.600	94.400	93.800	93.200	91.800	93.600	95.000	95.400	93.800	92.600	92.200	93.600
$\{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	True	16.038	1.031	0.315	0.193	0.103	0.049	16.038	1.031	0.315	0.193	0.103	0.049
$\{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	Median	16.008	1.029	0.325	0.204	0.112	0.057	16.010	1.009	0.321	0.202	0.113	0.059
$\{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	Bias	-0.030	-0.002	0.009	0.011	0.008	0.007	-0.028	-0.022	0.006	0.010	0.010	0.009
$\{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	Coverage	94.800	95.600	93.600	94.200	93.400	92.800	94.400	95.200	93.400	94.000	91.600	93.000
$\{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	True	31.710	7.136	3.098	2.016	1.474	1.081	31.710	7.136	3.098	2.016	1.474	1.081
$\{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	Median	31.675	7.124	3.134	2.065	1.514	1.120	31.684	7.040	3.087	2.034	1.497	1.118
$\{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	Bias	-0.035	-0.012	0.036	0.049	0.040	0.039	-0.026	-0.096	-0.011	0.018	0.023	0.037
$\{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	Coverage	93.400	96.000	94.800	94.200	94.000	93.200	94.200	95.000	94.800	95.000	94.800	93.600
$\{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	True	12.103	6.450	3.433	2.312	1.844	1.493	12.103	6.450	3.433	2.312	1.844	1.493
$\{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	Median	12.097	6.438	3.445	2.336	1.865	1.515	12.098	6.396	3.406	2.307	1.844	1.506
$\{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	Bias	-0.006	-0.012	0.013	0.023	0.021	0.022	-0.005	-0.054	-0.026	-0.005	0.001	0.013
$\{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	Coverage	95.200	96.600	96.000	95.200	94.800	93.200	95.200	94.400	95.000	94.400	94.800	93.600
$\{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	True	15.610	16.505	10.570	7.462	6.238	5.310	15.610	16.505	10.570	7.462	6.238	5.310
$\{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	Median	15.612	16.476	10.561	7.474	6.253	5.328	15.610	16.427	10.485	7.413	6.207	5.305
$\{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	Bias	0.002	-0.029	-0.009	0.013	0.015	0.019	0.001	-0.079	-0.085	-0.049	-0.032	-0.005
$\{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	Coverage	94.600	93.000	93.800	94.600	93.800	92.400	93.800	92.200	93.800	93.200	92.400	92.200
$\{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	True	9.229	23.965	19.630	15.200	13.219	11.687	9.229	23.965	19.630	15.200	13.219	11.687
$\{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	Median	9.249	23.934	19.561	15.134	13.162	11.635	9.244	23.924	19.502	15.078	13.122	11.620
$\{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	Bias	0.020	-0.032	-0.068	-0.065	-0.057	-0.052	0.015	-0.041	-0.128	-0.121	-0.097	-0.067
$\{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	Coverage	93.800	92.400	92.200	92.600	92.200	92.400	93.600	91.600	92.600	92.000	91.000	91.200
$\{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	True	2.880	24.354	29.626	30.016	27.051	24.708	2.880	24.354	29.626	30.016	27.051	24.708
$\{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	Median	2.907	24.393	29.689	30.079	27.091	24.724	2.902	24.430	29.704	30.089	27.109	24.743
$\{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	Bias	0.027	0.039	0.063	0.062	0.040	0.016	0.022	0.076	0.079	0.072	0.058	0.034
$\{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	Coverage	92.800	95.200	96.000	95.800	94.600	95.600	93.400	95.000	95.200	95.000	95.000	95.000

† Results from 500 replications with each dataset including 2000 individuals. The true marker evolution was based on linear splines with knots at 1 and 5 years since baseline and it was correctly specified in the fitted SREM-CIF-1 and SREM-CIF-2 models. "True" denotes the true parameter values; "Median" the mean of posterior medians over the 500 replications; "Bias" the mean bias for posterior median estimates; and "Coverage" the empirical coverage probability of posterior credible intervals.

Table 5.15: Simulation study results for marker state probabilities (%) for group 1 ($w = 1$) when the data have been generated by SREM-CIF-2 and there is misclassification†.

State	Est	$t=0$	$t=2$	$t=4$	$t=6$	$t=8$	$t=10$	$t=0$	$t=2$	$t=4$	$t=6$	$t=8$	$t=10$
		Results from SREM-CIF-1						Results from SREM-CIF-2					
$\{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	True	12.431	0.198	0.048	0.031	0.014	0.005	12.431	0.198	0.048	0.031	0.014	0.005
$\{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	Median	12.472	0.204	0.050	0.033	0.014	0.006	12.465	0.202	0.052	0.034	0.016	0.007
$\{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	Bias	0.041	0.005	0.002	0.002	0.001	0.001	0.035	0.004	0.004	0.004	0.002	0.002
$\{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	Coverage	94.600	95.000	95.200	95.200	92.400	92.200	94.000	95.200	94.800	94.800	92.600	93.200
$\{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	True	16.038	1.015	0.333	0.223	0.145	0.092	16.038	1.015	0.333	0.223	0.145	0.092
$\{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	Median	16.034	1.043	0.349	0.237	0.152	0.097	16.036	1.027	0.347	0.237	0.155	0.101
$\{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	Bias	-0.004	0.028	0.016	0.014	0.007	0.005	-0.002	0.012	0.015	0.014	0.009	0.008
$\{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	Coverage	94.400	93.400	96.000	94.000	93.400	92.800	95.000	95.400	95.600	94.800	92.800	93.200
$\{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	True	31.710	6.990	3.111	2.110	1.643	1.305	31.710	6.990	3.111	2.110	1.643	1.305
$\{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	Median	31.680	7.084	3.203	2.195	1.706	1.358	31.686	7.023	3.174	2.172	1.689	1.346
$\{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	Bias	-0.030	0.094	0.092	0.085	0.064	0.054	-0.023	0.033	0.063	0.062	0.046	0.042
$\{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	Coverage	95.200	93.000	93.600	92.200	91.800	93.000	95.200	94.200	94.600	94.400	92.200	92.800
$\{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	True	12.103	6.336	3.415	2.364	1.953	1.650	12.103	6.336	3.415	2.364	1.953	1.650
$\{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	Median	12.085	6.367	3.467	2.417	2.000	1.695	12.088	6.342	3.443	2.395	1.978	1.674
$\{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	Bias	-0.018	0.032	0.053	0.053	0.047	0.045	-0.015	0.007	0.029	0.032	0.025	0.024
$\{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	Coverage	94.200	93.000	92.000	91.600	91.200	91.400	94.000	94.000	93.000	92.800	92.800	92.200
$\{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	True	15.610	16.278	10.503	7.565	6.483	5.674	15.610	16.278	10.503	7.565	6.483	5.674
$\{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	Median	15.585	16.266	10.551	7.631	6.551	5.744	15.586	16.258	10.511	7.587	6.497	5.688
$\{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	Bias	-0.025	-0.012	0.048	0.066	0.068	0.070	-0.024	-0.020	0.008	0.022	0.014	0.015
$\{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	Coverage	95.200	93.600	93.400	92.800	91.200	91.800	95.000	93.800	93.800	93.800	92.400	93.200
$\{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	True	9.229	23.750	19.550	15.361	13.602	12.257	9.229	23.750	19.550	15.361	13.602	12.257
$\{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	Median	9.221	23.628	19.458	15.306	13.564	12.238	9.219	23.679	19.456	15.276	13.514	12.170
$\{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	Bias	-0.008	-0.122	-0.093	-0.056	-0.038	-0.019	-0.010	-0.071	-0.095	-0.086	-0.089	-0.087
$\{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	Coverage	94.200	94.400	93.800	93.200	93.400	93.400	94.600	94.400	93.600	92.600	92.600	93.000
$\{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	True	2.880	24.253	29.643	30.360	27.724	25.659	2.880	24.253	29.643	30.360	27.724	25.659
$\{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	Median	2.893	24.148	29.519	30.261	27.640	25.599	2.890	24.245	29.627	30.334	27.671	25.583
$\{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	Bias	0.013	-0.105	-0.123	-0.099	-0.084	-0.060	0.010	-0.008	-0.016	-0.026	-0.053	-0.076
$\{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	Coverage	94.800	93.400	94.400	94.400	94.000	95.200	95.400	93.800	94.800	94.600	94.400	95.400

† Results from 500 replications with each dataset including 2000 individuals. The true marker evolution was based on linear splines with knots at 1 and 5 years since baseline and it was correctly specified in the fitted SREM-CIF-1 and SREM-CIF-2 models. "True" denotes the true parameter values; "Median" the mean of posterior medians over the 500 replications; "Bias" the mean bias for posterior median estimates; and "Coverage" the empirical coverage probability of posterior credible intervals.

Table 5.16: Simulation study results for marker state probabilities (%) for group 0 ($w = 0$) when the data have been generated by SREM-CIF-2 and there is misclassification†.

State	Est	$t=0$	$t=2$	$t=4$	$t=6$	$t=8$	$t=10$	$t=0$	$t=2$	$t=4$	$t=6$	$t=8$	$t=10$
		Results from SREM-CIF-1						Results from SREM-CIF-2					
$\{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	True	12.431	0.198	0.048	0.031	0.014	0.005	12.431	0.198	0.048	0.031	0.014	0.005
$\{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	Median	12.472	0.204	0.050	0.033	0.014	0.006	12.465	0.202	0.052	0.034	0.016	0.007
$\{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	Bias	0.041	0.005	0.002	0.002	0.001	0.001	0.035	0.004	0.004	0.004	0.002	0.002
$\{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	Coverage	94.600	95.000	95.200	95.200	92.400	92.200	94.000	95.200	94.800	94.800	92.600	93.200
$\{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	True	16.038	1.015	0.333	0.223	0.145	0.092	16.038	1.015	0.333	0.223	0.145	0.092
$\{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	Median	16.034	1.043	0.349	0.237	0.152	0.097	16.036	1.027	0.347	0.237	0.155	0.101
$\{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	Bias	-0.004	0.028	0.016	0.014	0.007	0.005	-0.002	0.012	0.015	0.014	0.009	0.008
$\{\sqrt{50} \leq m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	Coverage	94.400	93.400	96.000	94.000	93.400	92.800	95.000	95.400	95.600	94.800	92.800	93.200
$\{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	True	31.710	6.990	3.111	2.110	1.643	1.305	31.710	6.990	3.111	2.110	1.643	1.305
$\{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	Median	31.680	7.084	3.203	2.195	1.706	1.358	31.686	7.023	3.174	2.172	1.689	1.346
$\{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	Bias	-0.030	0.094	0.092	0.085	0.064	0.054	-0.023	0.033	0.063	0.062	0.046	0.042
$\{\sqrt{100} \leq m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	Coverage	95.200	93.000	93.600	92.200	91.800	93.000	95.200	94.200	94.600	94.400	92.200	92.800
$\{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	True	12.103	6.336	3.415	2.364	1.953	1.650	12.103	6.336	3.415	2.364	1.953	1.650
$\{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	Median	12.085	6.367	3.467	2.417	2.000	1.695	12.088	6.342	3.443	2.395	1.978	1.674
$\{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	Bias	-0.018	0.032	0.053	0.053	0.047	0.045	-0.015	0.007	0.029	0.032	0.025	0.024
$\{\sqrt{200} \leq m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	Coverage	94.200	93.000	92.000	91.600	91.200	91.400	94.000	94.000	93.000	92.800	92.800	92.200
$\{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	True	15.610	16.278	10.503	7.565	6.483	5.674	15.610	16.278	10.503	7.565	6.483	5.674
$\{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	Median	15.585	16.266	10.551	7.631	6.551	5.744	15.586	16.258	10.511	7.587	6.497	5.688
$\{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	Bias	-0.025	-0.012	0.048	0.066	0.068	0.070	-0.024	-0.020	0.008	0.022	0.014	0.015
$\{\sqrt{250} \leq m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	Coverage	95.200	93.600	93.400	92.800	91.200	91.800	95.000	93.800	93.800	93.800	92.400	93.200
$\{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	True	9.229	23.750	19.550	15.361	13.602	12.257	9.229	23.750	19.550	15.361	13.602	12.257
$\{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	Median	9.221	23.628	19.458	15.306	13.564	12.238	9.219	23.679	19.456	15.276	13.514	12.170
$\{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	Bias	-0.008	-0.122	-0.093	-0.056	-0.038	-0.019	-0.010	-0.071	-0.095	-0.086	-0.089	-0.087
$\{\sqrt{350} \leq m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	Coverage	94.200	94.400	93.800	93.200	93.400	93.400	94.600	94.400	93.600	92.600	92.600	93.000
$\{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	True	2.880	24.253	29.643	30.360	27.724	25.659	2.880	24.253	29.643	30.360	27.724	25.659
$\{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	Median	2.893	24.148	29.519	30.261	27.640	25.599	2.890	24.245	29.627	30.334	27.671	25.583
$\{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	Bias	0.013	-0.105	-0.123	-0.099	-0.084	-0.060	0.010	-0.008	-0.016	-0.026	-0.053	-0.076
$\{\sqrt{500} \leq m_i(t)\} \cap T_i^\star > t$	Coverage	94.800	93.400	94.400	94.400	94.000	95.200	95.400	93.800	94.800	94.600	94.400	95.400

† Results from 500 replications with each dataset including 2000 individuals. The true marker evolution was based on linear splines with knots at 1 and 5 years since baseline and it was correctly specified in the fitted SREM-CIF-1 and SREM-CIF-2 models. "True" denotes the true parameter values; "Median" the mean of posterior medians over the 500 replications; "Bias" the mean bias for posterior median estimates; and "Coverage" the empirical coverage probability of posterior credible intervals.

5.6.2.2 Simulation study results for transition marker state probabilities by baseline marker state under misclassified failure cause

The transition marker probabilities by baseline marker state are shown in Tables 5.17-5.20. Both models had great performance leading to estimates with small biases and acceptable coverage probabilities in almost all cases. However, when the data were generated from the SREM-CIF-1 model and the SREM-CIF-2 model was fitted, the coverage probability for the "$m_i(0) < \sqrt{50} \rightarrow \{m_i(6) < \sqrt{50}\} \cap (T_i^\star > t)$" transition was equal to 82.8% (Table 5.17). Recall that the same finding was observed in the simulation study with known failure cause (Table 5.7).

Table 5.17: Simulation study results for transition marker state probabilities (%) by baseline marker state for group 1 ($w = 1$) when the data have been generated by SREM-CIF-1 and there is misclassification†.

transition	$t=2$	$t=4$	$t=6$	$t=8$	$t=10$	$t=2$	$t=4$	$t=6$	$t=8$	$t=10$
	Results from SREM-CIF-1					Results from SREM-CIF-2				
$m_i(0) < \sqrt{50} \to \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	1.31/1.31/94.6	0.15/0.17/93.2	0.03/0.04/93.4	<0.01/0.01/92.6	<0.01/<0.01/94.2	1.31/1.33/94.6	0.15/0.19/89.4	0.03/0.06/82.8	<0.01/0.01/85.4	<0.01/<0.01/90.6
$m_i(0) < \sqrt{50} \to \{\sqrt{50} \le m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	5.86/5.85/95.4	1.09/1.13/92.8	0.36/0.4/93.8	0.13/0.16/93.4	0.03/0.05/93.2	5.86/5.79/94.6	1.09/1.17/91.2	0.36/0.45/86.8	0.13/0.21/85.4	0.03/0.08/86.6
$m_i(0) < \sqrt{50} \to \{\sqrt{100} \le m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	24.64/24.57/95.4	7.77/7.88/93.4	3.44/3.56/93	2.25/2.36/92.8	1.42/1.53/94	24.64/24.27/93.6	7.77/7.89/94	3.44/3.68/90.6	2.25/2.53/89.4	1.42/1.71/89.6
$m_i(0) < \sqrt{50} \to \{\sqrt{200} \le m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	12.55/12.5/95	6.45/6.47/95.6	3.4/3.44/93.8	2.57/2.62/94.2	1.95/2/95.2	12.55/12.37/92.8	6.45/6.43/95	3.4/3.48/93.4	2.57/2.68/93.4	1.95/2.08/92.2
$m_i(0) < \sqrt{50} \to \{\sqrt{250} \le m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	15.7/15.64/93.6	14.04/14/95.8	8.96/8.97/94.4	7.26/7.29/94.8	5.96/6/95.2	15.7/15.49/91.2	14.04/13.88/95.4	8.96/8.98/95	7.26/7.35/94.8	5.96/6.09/95.2
$m_i(0) < \sqrt{50} \to \{\sqrt{350} \le m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	6.3/6.29/93.6	14.75/14.64/93	13.06/12.96/95.6	11.18/11.1/95	9.71/9.65/95.4	6.3/6.24/92.2	14.75/14.52/92.6	13.06/12.91/94.2	11.18/11.09/94.6	9.71/9.66/94.8
$m_i(0) < \sqrt{50} \to \{\sqrt{500} \le m_i(t)\} \cap T_i^\star > t$	0.66/0.67/94.4	7.56/7.52/93.6	12.77/12.69/91.8	11.4/11.33/92.6	10.31/10.23/93.4	0.66/0.67/93.6	7.56/7.47/92	12.77/12.63/92.8	11.4/11.28/93.2	10.31/10.19/93.2
$\sqrt{50} \le m_i(0) < \sqrt{100} \to \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	0.11/0.11/94.2	0.06/0.07/93.2	0.03/0.04/93.8	0.01/0.01/92.8	<0.01/<0.01/93.2	0.11/0.11/94.4	0.06/0.07/92.8	0.03/0.04/92.6	0.01/0.01/89.6	<0.01/<0.01/91.2
$\sqrt{50} \le m_i(0) < \sqrt{100} \to \{\sqrt{50} \le m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	1.27/1.28/95.2	0.52/0.54/93.6	0.31/0.33/94.6	0.17/0.18/92.8	0.07/0.09/92.6	1.27/1.26/93.6	0.52/0.54/93.8	0.31/0.33/93.6	0.17/0.19/91.6	0.07/0.1/90.8
$\sqrt{50} \le m_i(0) < \sqrt{100} \to \{\sqrt{100} \le m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	13.93/13.95/94.4	5.2/5.29/93.6	3.03/3.12/93.2	2.23/2.31/92.2	1.64/1.72/93.2	13.93/13.86/95.2	5.2/5.23/94.2	3.03/3.09/94	2.23/2.29/93.4	1.64/1.72/93.6
$\sqrt{50} \le m_i(0) < \sqrt{100} \to \{\sqrt{200} \le m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	12.83/12.83/96	5.45/5.5/94	3.24/3.29/93.4	2.6/2.65/93.2	2.11/2.17/94	12.83/12.78/95.4	5.45/5.45/95	3.24/3.26/93.6	2.6/2.62/93.8	2.11/2.15/95.6
$\sqrt{50} \le m_i(0) < \sqrt{100} \to \{\sqrt{250} \le m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	25.61/25.59/94.4	14.79/14.84/96.2	9.5/9.57/94	7.98/8.05/95	6.82/6.89/95	25.61/25.52/93.8	14.79/14.75/94.8	9.5/9.49/94.8	7.98/7.98/96	6.82/6.83/95.6
$\sqrt{50} \le m_i(0) < \sqrt{100} \to \{\sqrt{350} \le m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	18.56/18.52/93.4	20.82/20.78/94.4	16.27/16.24/94.4	14.21/14.2/94	12.61/12.61/95.4	18.56/18.48/93	20.82/20.71/94.8	16.27/16.17/94.6	14.21/14.13/94.8	12.61/12.53/95.6
$\sqrt{50} \le m_i(0) < \sqrt{100} \to \{\sqrt{500} \le m_i(t)\} \cap T_i^\star > t$	3.83/3.84/94	15.7/15.67/93.2	20.71/20.68/92.4	18.71/18.67/92	17.11/17.07/92.4	3.83/3.83/94.4	15.7/15.65/93.6	20.71/20.66/93.6	18.71/18.64/92.4	17.11/17.02/92.4
$\sqrt{100} \le m_i(0) < \sqrt{200} \to \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	0.01/0.01/94.8	0.02/0.02/93.2	0.02/0.02/94.2	0.01/0.01/93.8	<0.01/<0.01/93.4	0.01/0.01/94.6	0.02/0.02/93.4	0.02/0.02/93.8	0.01/0.01/93	<0.01/<0.01/93
$\sqrt{100} \le m_i(0) < \sqrt{200} \to \{\sqrt{50} \le m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	0.22/0.23/94.8	0.22/0.23/93.2	0.2/0.21/93.6	0.12/0.13/93.8	0.07/0.08/92	0.22/0.22/95.2	0.22/0.22/94	0.2/0.21/94	0.12/0.13/93.2	0.07/0.08/93.2
$\sqrt{100} \le m_i(0) < \sqrt{200} \to \{\sqrt{100} \le m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	4.96/4.97/95.6	2.93/2.97/94.4	2.19/2.26/94	1.68/1.74/93.8	1.3/1.36/92.4	4.96/4.96/96.2	2.93/2.94/94.2	2.19/2.21/94.8	1.68/1.69/94.2	1.3/1.32/94.2
$\sqrt{100} \le m_i(0) < \sqrt{200} \to \{\sqrt{200} \le m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	7.19/7.19/95.4	3.73/3.76/95.2	2.6/2.64/94	2.13/2.17/92.6	1.78/1.82/94.2	7.19/7.19/95.6	3.73/3.73/94.8	2.6/2.6/94.6	2.13/2.13/94.4	1.78/1.78/94.2
$\sqrt{100} \le m_i(0) < \sqrt{200} \to \{\sqrt{250} \le m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	22.83/22.84/94.4	12.37/12.41/95.6	8.53/8.58/94.8	7.27/7.33/94	6.31/6.38/95	22.83/22.88/94.2	12.37/12.37/96	8.53/8.52/94.4	7.27/7.24/94.8	6.31/6.28/95.4
$\sqrt{100} \le m_i(0) < \sqrt{200} \to \{\sqrt{350} \le m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	31.24/31.23/94.4	23.02/23.02/94.4	17.15/17.15/93.4	15.12/15.13/94.4	13.55/13.56/94.4	31.24/31.29/94.4	23.02/23.02/95.2	17.15/17.11/93.2	15.12/15.06/94.2	13.55/13.47/95.2
$\sqrt{100} \le m_i(0) < \sqrt{200} \to \{\sqrt{500} \le m_i(t)\} \cap T_i^\star > t$	14.48/14.48/94	26.38/26.41/93.8	28.92/28.97/93.4	26.31/26.34/93.4	24.23/24.25/93	14.48/14.5/93.8	26.38/26.47/94.4	28.92/29.03/93	26.31/26.36/93.4	24.23/24.21/92.8
$\sqrt{200} \le m_i(0) < \sqrt{250} \to \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	<0.01/<0.01/95.4	0.01/0.01/93.4	0.01/0.01/94.2	0.01/0.01/93	<0.01/<0.01/93.4	<0.01/<0.01/95.6	0.01/0.01/93.2	0.01/0.01/94.6	0.01/0.01/95.6	<0.01/<0.01/94.4
$\sqrt{200} \le m_i(0) < \sqrt{250} \to \{\sqrt{50} \le m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	0.03/0.03/95.6	0.09/0.09/93.8	0.12/0.13/94.2	0.08/0.08/93.6	0.05/0.05/93.6	0.03/0.03/95.8	0.09/0.09/93.2	0.12/0.12/95	0.08/0.08/93.8	0.05/0.05/95.2
$\sqrt{200} \le m_i(0) < \sqrt{250} \to \{\sqrt{100} \le m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	1.32/1.33/96	1.56/1.58/94.8	1.53/1.57/95	1.2/1.24/94	0.96/0.99/93.6	1.32/1.33/95.6	1.56/1.57/94.8	1.53/1.54/94.2	1.2/1.19/94.6	0.96/0.95/94.2
$\sqrt{200} \le m_i(0) < \sqrt{250} \to \{\sqrt{200} \le m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	2.94/2.94/95	2.37/2.39/96	2/2.02/94.8	1.66/1.68/94.2	1.4/1.43/93.2	2.94/2.95/95	2.37/2.37/95.2	2/2/95.2	1.66/1.64/93	1.4/1.39/95
$\sqrt{200} \le m_i(0) < \sqrt{250} \to \{\sqrt{250} \le m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	14.64/14.64/95.4	9.45/9.47/94.8	7.26/7.29/94.6	6.24/6.27/93.8	5.46/5.51/94.4	14.64/14.69/95	9.45/9.46/95.2	7.26/7.24/94.6	6.24/6.19/94	5.46/5.41/93.8
$\sqrt{200} \le m_i(0) < \sqrt{250} \to \{\sqrt{350} \le m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	34.65/34.65/94.4	22.37/22.36/93.6	16.78/16.76/93.4	14.88/14.87/95.2	13.4/13.4/94.2	34.65/34.79/93.6	22.37/22.41/93.8	16.78/16.74/93	14.88/14.8/94.6	13.4/13.3/94.2
$\sqrt{200} \le m_i(0) < \sqrt{250} \to \{\sqrt{500} \le m_i(t)\} \cap T_i^\star > t$	30.23/30.24/95.6	36.64/36.74/94.4	36.1/36.22/94.4	32.96/33.07/93.2	30.46/30.55/93.2	30.23/30.36/96.4	36.64/36.9/94.2	36.1/36.34/93.8	32.96/33.11/93	30.46/30.51/93.2
$\sqrt{250} \le m_i(0) < \sqrt{350} \to \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	<0.01/<0.01/95.6	<0.01/<0.01/93.6	0.01/0.01/94.8	<0.01/<0.01/94	<0.01/<0.01/92.6	<0.01/<0.01/95.2	<0.01/<0.01/93	0.01/0.01/94.6	<0.01/<0.01/95.2	<0.01/<0.01/95.2
$\sqrt{250} \le m_i(0) < \sqrt{350} \to \{\sqrt{50} \le m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	0.01/0.01/95.8	0.04/0.04/94.6	0.08/0.08/94.4	0.05/0.06/94.8	0.04/0.04/94.2	0.01/0.01/95	0.04/0.04/94	0.08/0.08/93.8	0.05/0.05/94.6	0.04/0.04/95.2
$\sqrt{250} \le m_i(0) < \sqrt{350} \to \{\sqrt{100} \le m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	0.43/0.44/96	0.91/0.93/95.6	1.12/1.14/94.2	0.88/0.91/94.6	0.71/0.74/94.2	0.43/0.44/95.4	0.91/0.92/95	1.12/1.12/94.6	0.88/0.87/94.2	0.71/0.7/95
$\sqrt{250} \le m_i(0) < \sqrt{350} \to \{\sqrt{200} \le m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	1.22/1.23/95.8	1.56/1.57/95.8	1.56/1.58/94.6	1.3/1.32/95.4	1.11/1.13/94.2	1.22/1.23/95.8	1.56/1.56/95.4	1.56/1.56/94.6	1.3/1.29/93.8	1.11/1.09/95.4
$\sqrt{250} \le m_i(0) < \sqrt{350} \to \{\sqrt{250} \le m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	8.27/8.27/94.6	7.11/7.11/94.8	6.12/6.14/94	5.29/5.31/95.4	4.65/4.68/95.2	8.27/8.3/94.8	7.11/7.11/95.2	6.12/6.1/95	5.29/5.23/93.8	4.65/4.59/94.8
$\sqrt{250} \le m_i(0) < \sqrt{350} \to \{\sqrt{350} \le m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	29.51/29.48/94.8	20.18/20.14/93.4	15.77/15.72/92.2	14.02/13.99/94.2	12.67/12.65/93.8	29.51/29.63/95.4	20.18/20.19/94.4	15.77/15.71/93.4	14.02/13.92/93	12.67/12.53/92.6
$\sqrt{250} \le m_i(0) < \sqrt{350} \to \{\sqrt{500} \le m_i(t)\} \cap T_i^\star > t$	46.01/46.04/95.2	44.93/45.08/94.4	41.65/41.83/95.4	38.13/38.29/93.4	35.34/35.46/93.8	46.01/46.26/95.4	44.93/45.31/94.4	41.65/41.99/94.2	38.13/38.34/94.2	35.34/35.41/94.8
$\sqrt{350} \le m_i(0) < \sqrt{500} \to \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	<0.01/<0.01/95.6	<0.01/<0.01/94.2	<0.01/<0.01/93.8	<0.01/<0.01/94	<0.01/<0.01/93.8	<0.01/<0.01/94.8	<0.01/<0.01/93.6	<0.01/<0.01/94	<0.01/<0.01/95.4	<0.01/<0.01/95.8
$\sqrt{350} \le m_i(0) < \sqrt{500} \to \{\sqrt{50} \le m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	<0.01/<0.01/95.8	0.01/0.01/94	0.04/0.04/94.2	0.03/0.03/94	0.02/0.02/94.4	<0.01/<0.01/95.2	0.01/0.01/94	0.04/0.04/93.6	0.03/0.03/94.2	0.02/0.02/94.4
$\sqrt{350} \le m_i(0) < \sqrt{500} \to \{\sqrt{100} \le m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	0.06/0.07/95.4	0.38/0.39/94	0.67/0.69/93.8	0.54/0.55/94.2	0.44/0.45/95.2	0.06/0.07/95.4	0.38/0.39/94.2	0.67/0.67/94.8	0.54/0.53/93.6	0.44/0.43/94.4
$\sqrt{350} \le m_i(0) < \sqrt{500} \to \{\sqrt{200} \le m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	0.26/0.26/94.8	0.78/0.78/95.2	1.04/1.05/94.4	0.87/0.88/93.8	0.75/0.76/94.8	0.26/0.26/94.4	0.78/0.78/94.8	1.04/1.03/94.6	0.87/0.86/94	0.75/0.73/94.6
$\sqrt{350} \le m_i(0) < \sqrt{500} \to \{\sqrt{250} \le m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	2.72/2.73/93.6	4.29/4.29/94.6	4.56/4.55/94	3.95/3.96/94.6	3.49/3.5/95.8	2.72/2.74/94.6	4.29/4.29/95.2	4.56/4.52/94.4	3.95/3.9/93.4	3.49/3.43/94
$\sqrt{350} \le m_i(0) < \sqrt{500} \to \{\sqrt{350} \le m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	17.44/17.41/94.4	15.79/15.7/95.4	13.69/13.6/93.8	12.21/12.14/94.2	11.07/11.01/93.8	17.44/17.5/94	15.79/15.76/95.2	13.69/13.6/93.8	12.21/12.07/93.8	11.07/10.9/93.2
$\sqrt{350} \le m_i(0) < \sqrt{500} \to \{\sqrt{500} \le m_i(t)\} \cap T_i^\star > t$	66.74/66.76/94.4	56.02/56.2/95.2	49.19/49.4/94.4	45.17/45.36/95.4	41.98/42.13/95.8	66.74/67.15/93.8	56.02/56.53/94.4	49.19/49.61/94	45.17/45.42/95	41.98/42.05/94.8
$\sqrt{500} \le m_i(0) \to \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	<0.01/<0.01/95.4	<0.01/<0.01/94	<0.01/<0.01/93.2	<0.01/<0.01/94.2	<0.01/<0.01/94.8	<0.01/<0.01/94.8	<0.01/<0.01/93.8	<0.01/<0.01/94.2	<0.01/<0.01/93.2	<0.01/<0.01/94.4
$\sqrt{500} \le m_i(0) \to \{\sqrt{50} \le m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	<0.01/<0.01/95	<0.01/<0.01/93.4	0.02/0.02/94	0.01/0.01/93.4	0.01/0.01/94.6	<0.01/<0.01/94.6	<0.01/<0.01/94.4	0.02/0.02/94	0.01/0.01/92.8	0.01/0.01/93.8
$\sqrt{500} \le m_i(0) \to \{\sqrt{100} \le m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	<0.01/<0.01/94.4	0.11/0.11/93.6	0.31/0.32/93.8	0.25/0.26/93.6	0.21/0.22/94.6	<0.01/<0.01/94.2	0.11/0.11/94.4	0.31/0.32/94	0.25/0.25/93.2	0.21/0.21/93
$\sqrt{500} \le m_i(0) \to \{\sqrt{200} \le m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	0.02/0.02/93.6	0.27/0.27/94.2	0.55/0.56/94.4	0.47/0.47/93.8	0.4/0.41/94.4	0.02/0.02/94	0.27/0.27/94.4	0.55/0.55/94.6	0.47/0.46/94.2	0.4/0.39/93.6
$\sqrt{500} \le m_i(0) \to \{\sqrt{250} \le m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	0.4/0.41/93.4	1.88/1.87/94.6	2.81/2.8/94.2	2.45/2.45/93.8	2.17/2.18/94.6	0.4/0.42/94	1.88/1.88/94.6	2.81/2.78/94.2	2.45/2.4/93.4	2.17/2.12/93.6
$\sqrt{500} \le m_i(0) \to \{\sqrt{350} \le m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	5.36/5.38/93.4	9.59/9.49/94.6	10.33/10.22/94	9.24/9.15/94	8.4/8.32/93.8	5.36/5.4/93.8	9.59/9.53/95.2	10.33/10.2/94.4	9.24/9.08/94	8.4/8.21/92.6
$\sqrt{500} \le m_i(0) \to \{\sqrt{500} \le m_i(t)\} \cap T_i^\star > t$	83.1/83.06/94	67.94/68.07/93.8	58.09/58.29/95.4	53.53/53.71/95.6	49.9/50.05/95	83.1/83.56/91	67.94/68.52/92.6	58.09/58.55/94.2	53.53/53.76/94.6	49.9/49.91/95.2

† Results from 500 replications with each dataset including 2000 individuals. The true marker evolution was based on linear splines with knots at 1 and 5 years since baseline and it was correctly specified in the fitted SREM-CIF-1 and SREM-CIF-2 models. The true parameter values, the mean of posterior medians over the 500 replications and the empirical coverage probability of posterior credible intervals are presented, respectively.

Table 5.18: Simulation study results for transition marker state probabilities (%) by baseline marker state for group 0 ($w = 0$) when the data have been generated by SREM-CIF-1 and there is misclassification†.

transition	$t=2$	$t=4$	$t=6$	$t=8$	$t=10$	$t=2$	$t=4$	$t=6$	$t=8$	$t=10$
	Results from SREM-CIF-1					Results from SREM-CIF2				
$m_i(0) < \sqrt{50} \to \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	1.38/1.38/95	0.17/0.18/93.6	0.03/0.04/92.2	<0.01/0.01/92.2	<0.01/<0.01/94	1.38/1.36/95.4	0.17/0.19/92.4	0.03/0.05/89	<0.01/0.01/88.4	<0.01/<0.01/92.4
$m_i(0) < \sqrt{50} \to \{\sqrt{50} \le m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	6.02/6/95.6	1.14/1.17/94	0.38/0.41/92.4	0.14/0.16/92.2	0.03/0.05/92.2	6.02/5.87/95.2	1.14/1.18/93.8	0.38/0.44/91.4	0.14/0.19/89	0.03/0.07/87.8
$m_i(0) < \sqrt{50} \to \{\sqrt{100} \le m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	24.91/24.8/94.6	7.9/7.96/95	3.48/3.57/93.2	2.25/2.33/93	1.39/1.47/92.6	24.91/24.38/92.2	7.9/7.86/94.6	3.48/3.61/94	2.25/2.43/92	1.39/1.61/90
$m_i(0) < \sqrt{50} \to \{\sqrt{200} \le m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	12.6/12.52/93.6	6.47/6.47/94.2	3.39/3.41/93.4	2.53/2.55/94.6	1.88/1.91/94.2	12.6/12.35/89.8	6.47/6.37/94.2	3.39/3.41/94.2	2.53/2.59/93.6	1.88/1.98/90.8
$m_i(0) < \sqrt{50} \to \{\sqrt{250} \le m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	15.69/15.61/93.6	13.99/13.9/94.6	8.85/8.83/93.6	7.07/7.06/94	5.72/5.71/94	15.69/15.42/91.8	13.99/13.71/94	8.85/8.77/94.6	7.07/7.09/94.4	5.72/5.82/93.4
$m_i(0) < \sqrt{50} \to \{\sqrt{350} \le m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	6.27/6.26/93.4	14.6/14.45/93.2	12.8/12.66/92.6	10.81/10.69/93.4	9.27/9.15/94	6.27/6.19/92.8	14.6/14.28/91	12.8/12.57/92	10.81/10.68/93.6	9.27/9.21/94.8
$m_i(0) < \sqrt{50} \to \{\sqrt{500} \le m_i(t)\} \cap T_i^\star > t$	0.66/0.67/94.6	7.44/7.39/92.8	12.44/12.34/94.8	10.96/10.86/93	9.8/9.69/93.4	0.66/0.66/94.2	7.44/7.31/92.4	12.44/12.25/93.6	10.96/10.83/93.8	9.8/9.69/93.4
$\sqrt{50} \le m_i(0) < \sqrt{100} \to \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	0.11/0.11/94.6	0.06/0.07/94.8	0.04/0.04/92.4	0.01/0.01/92	<0.01/<0.01/93.6	0.11/0.11/95.2	0.06/0.07/93.8	0.04/0.04/92.4	0.01/0.01/91.2	<0.01/<0.01/92.4
$\sqrt{50} \le m_i(0) < \sqrt{100} \to \{\sqrt{50} \le m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	1.29/1.29/95.4	0.53/0.55/94.2	0.32/0.33/92.8	0.17/0.18/92.2	0.07/0.08/92.6	1.29/1.27/94.6	0.53/0.53/93.8	0.32/0.33/93.8	0.17/0.18/91.8	0.07/0.09/90.8
$\sqrt{50} \le m_i(0) < \sqrt{100} \to \{\sqrt{100} \le m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	13.95/13.96/94.2	5.19/5.26/94	3/3.07/94.2	2.18/2.24/93.6	1.58/1.63/93.8	13.95/13.82/93.8	5.19/5.17/94.4	3/3.02/93.6	2.18/2.21/94.2	1.58/1.64/93.4
$\sqrt{50} \le m_i(0) < \sqrt{100} \to \{\sqrt{200} \le m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	12.8/12.79/94.4	5.4/5.43/94.6	3.18/3.22/93.6	2.52/2.55/93.8	2.02/2.05/93.6	12.8/12.7/93.6	5.4/5.36/94.4	3.18/3.17/93.6	2.52/2.52/94	2.02/2.04/94.4
$\sqrt{50} \le m_i(0) < \sqrt{100} \to \{\sqrt{250} \le m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	25.47/25.42/95	14.59/14.6/94.6	9.27/9.3/93.2	7.68/7.72/93	6.48/6.51/92.8	25.47/25.28/93.8	14.59/14.46/94.8	9.27/9.21/93	7.68/7.66/94	6.48/6.5/93.4
$\sqrt{50} \le m_i(0) < \sqrt{100} \to \{\sqrt{350} \le m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	18.41/18.36/95.4	20.45/20.37/93.6	15.79/15.73/93.2	13.62/13.56/92.6	11.93/11.89/93.6	18.41/18.27/94.2	20.45/20.24/93	15.79/15.64/93.6	13.62/13.52/93.2	11.93/11.89/93.6
$\sqrt{50} \le m_i(0) < \sqrt{100} \to \{\sqrt{500} \le m_i(t)\} \cap T_i^\star > t$	3.79/3.8/95.2	15.36/15.32/94.2	20.02/19.96/94.4	17.87/17.8/94.6	16.17/16.1/95	3.79/3.78/94.4	15.36/15.25/94.2	20.02/19.92/94.6	17.87/17.8/94.6	16.17/16.12/95.2
$\sqrt{100} \le m_i(0) < \sqrt{200} \to \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	0.01/0.01/95.6	0.02/0.02/94.4	0.02/0.03/94.6	0.01/0.01/91.4	<0.01/<0.01/93.2	0.01/0.01/95.6	0.02/0.02/94.2	0.02/0.02/94	0.01/0.01/93.2	<0.01/<0.01/93.8
$\sqrt{100} \le m_i(0) < \sqrt{200} \to \{\sqrt{50} \le m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	0.23/0.23/95.6	0.22/0.23/94.8	0.2/0.21/94.6	0.12/0.13/93.8	0.07/0.07/93	0.23/0.22/95.4	0.22/0.22/95.2	0.2/0.2/94.6	0.12/0.12/95	0.07/0.07/93.6
$\sqrt{100} \le m_i(0) < \sqrt{200} \to \{\sqrt{100} \le m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	4.95/4.95/96	2.9/2.94/96.6	2.15/2.2/94.8	1.63/1.67/95	1.25/1.28/94	4.95/4.93/96.8	2.9/2.89/97	2.15/2.15/95.2	1.63/1.63/95.6	1.25/1.25/94.6
$\sqrt{100} \le m_i(0) < \sqrt{200} \to \{\sqrt{200} \le m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	7.14/7.14/96.2	3.67/3.69/96.4	2.53/2.56/95	2.05/2.07/94.2	1.69/1.71/93	7.14/7.12/95.8	3.67/3.65/95.6	2.53/2.52/95	2.05/2.04/94.6	1.69/1.69/93.8
$\sqrt{100} \le m_i(0) < \sqrt{200} \to \{\sqrt{250} \le m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	22.64/22.63/93.8	12.14/12.16/94.6	8.27/8.3/94.2	6.96/6.99/93.6	5.96/6/93.4	22.64/22.61/93	12.14/12.09/93.8	8.27/8.23/94.4	6.96/6.93/94.6	5.96/5.95/93.2
$\sqrt{100} \le m_i(0) < \sqrt{200} \to \{\sqrt{350} \le m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	30.91/30.88/95.2	22.51/22.48/94.6	16.56/16.53/93.6	14.43/14.4/93.8	12.77/12.75/93.6	30.91/30.86/95.4	22.51/22.42/94.4	16.56/16.47/93.6	14.43/14.36/93	12.77/12.73/93.4
$\sqrt{100} \le m_i(0) < \sqrt{200} \to \{\sqrt{500} \le m_i(t)\} \cap T_i^\star > t$	14.31/14.3/95	25.73/25.74/94.8	27.87/27.89/94.2	25.05/25.05/94.8	22.83/22.81/96.4	14.31/14.29/95	25.73/25.73/94.6	27.87/27.89/94.4	25.05/25.08/95.2	22.83/22.85/96.2
$\sqrt{200} \le m_i(0) < \sqrt{250} \to \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	<0.01/<0.01/95.8	0.01/0.01/94.2	0.01/0.01/95.2	0.01/0.01/93.6	<0.01/<0.01/93.6	<0.01/<0.01/96	0.01/0.01/94	0.01/0.01/95.2	0.01/0.01/93.8	<0.01/<0.01/93.4
$\sqrt{200} \le m_i(0) < \sqrt{250} \to \{\sqrt{50} \le m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	0.03/0.03/95.8	0.09/0.09/95.2	0.12/0.13/96	0.08/0.08/94.6	0.05/0.05/94.2	0.03/0.03/95.8	0.09/0.09/94.6	0.12/0.12/95.6	0.08/0.08/94.4	0.05/0.05/95
$\sqrt{200} \le m_i(0) < \sqrt{250} \to \{\sqrt{100} \le m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	1.32/1.32/96.6	1.54/1.56/95.8	1.5/1.53/95.8	1.16/1.18/95.6	0.91/0.94/94.6	1.32/1.32/95.6	1.54/1.54/95.2	1.5/1.49/96.2	1.16/1.14/95.2	0.91/0.9/94.4
$\sqrt{200} \le m_i(0) < \sqrt{250} \to \{\sqrt{200} \le m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	2.92/2.92/95.6	2.33/2.34/95.6	1.94/1.96/96	1.59/1.6/95	1.33/1.34/93.8	2.92/2.92/95.6	2.33/2.32/95	1.94/1.93/96.2	1.59/1.57/95.4	1.33/1.32/94.8
$\sqrt{200} \le m_i(0) < \sqrt{250} \to \{\sqrt{250} \le m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	14.49/14.48/96.6	9.24/9.25/94.8	7.01/7.03/95	5.95/5.97/94.4	5.15/5.17/94.6	14.49/14.49/96	9.24/9.21/95	7.01/6.97/94.2	5.95/5.91/94.8	5.15/5.12/94
$\sqrt{200} \le m_i(0) < \sqrt{250} \to \{\sqrt{350} \le m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	34.23/34.22/94	21.83/21.79/93.2	16.16/16.12/92.8	14.16/14.12/92.2	12.61/12.57/93.4	34.23/34.27/94.4	21.83/21.78/93.6	16.16/16.08/93	14.16/14.08/93.6	12.61/12.54/93.8
$\sqrt{200} \le m_i(0) < \sqrt{250} \to \{\sqrt{500} \le m_i(t)\} \cap T_i^\star > t$	29.84/29.84/95.6	35.69/35.77/96.8	34.72/34.82/96	31.35/31.41/95.6	28.67/28.72/95	29.84/29.88/96.6	35.69/35.82/96	34.72/34.86/95.4	31.35/31.46/95.6	28.67/28.76/95.4
$\sqrt{250} \le m_i(0) < \sqrt{350} \to \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	<0.01/<0.01/95.2	<0.01/<0.01/94.8	0.01/0.01/95.2	<0.01/<0.01/93	<0.01/<0.01/94	<0.01/<0.01/95.2	<0.01/<0.01/93.4	0.01/0.01/95	<0.01/<0.01/93.8	<0.01/<0.01/93
$\sqrt{250} \le m_i(0) < \sqrt{350} \to \{\sqrt{50} \le m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	0.01/0.01/95.4	0.04/0.04/95.2	0.08/0.08/95.4	0.05/0.06/95.4	0.03/0.04/94.6	0.01/0.01/95.4	0.04/0.04/94	0.08/0.08/95.2	0.05/0.05/94	0.03/0.03/94.8
$\sqrt{250} \le m_i(0) < \sqrt{350} \to \{\sqrt{100} \le m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	0.43/0.43/95.8	0.9/0.91/95.8	1.08/1.11/95.2	0.85/0.87/95	0.68/0.69/95.2	0.43/0.43/95	0.9/0.9/94.6	1.08/1.08/95.2	0.85/0.84/95	0.68/0.67/95.2
$\sqrt{250} \le m_i(0) < \sqrt{350} \to \{\sqrt{200} \le m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	1.21/1.22/95	1.53/1.53/96	1.51/1.52/94.8	1.25/1.26/95.6	1.05/1.06/96	1.21/1.22/95.6	1.53/1.53/94.6	1.51/1.5/95.2	1.25/1.23/95.4	1.05/1.03/95.4
$\sqrt{250} \le m_i(0) < \sqrt{350} \to \{\sqrt{250} \le m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	8.18/8.17/94.2	6.94/6.93/94.6	5.91/5.91/94.8	5.04/5.04/94.4	4.38/4.39/94.8	8.18/8.19/95	6.94/6.92/93.6	5.91/5.86/94	5.04/4.98/94.6	4.38/4.33/94.4
$\sqrt{250} \le m_i(0) < \sqrt{350} \to \{\sqrt{350} \le m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	29.14/29.1/94.8	19.67/19.61/92.8	15.17/15.11/93.4	13.33/13.27/92.4	11.91/11.85/92.6	29.14/29.16/95.2	19.67/19.61/93.6	15.17/15.08/92.6	13.33/13.23/92.6	11.91/11.81/92.2
$\sqrt{250} \le m_i(0) < \sqrt{350} \to \{\sqrt{500} \le m_i(t)\} \cap T_i^\star > t$	45.38/45.4/95.6	43.73/43.86/96	40.04/40.19/95.4	36.24/36.36/95.8	33.25/33.32/96.2	45.38/45.51/95.6	43.73/43.96/96.6	40.04/40.26/95	36.24/36.4/95.6	33.25/33.35/96
$\sqrt{350} \le m_i(0) < \sqrt{500} \to \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	<0.01/<0.01/95.4	<0.01/<0.01/94.4	<0.01/<0.01/94.6	<0.01/<0.01/94.2	<0.01/<0.01/94.2	<0.01/<0.01/95	<0.01/<0.01/93.6	<0.01/<0.01/94.6	<0.01/<0.01/94	<0.01/<0.01/93.2
$\sqrt{350} \le m_i(0) < \sqrt{500} \to \{\sqrt{50} \le m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	<0.01/<0.01/95.4	0.01/0.01/94.6	0.04/0.04/94.8	0.03/0.03/95.2	0.02/0.02/94.6	<0.01/<0.01/95.2	0.01/0.01/94.4	0.04/0.04/95	0.03/0.03/93.4	0.02/0.02/93.6
$\sqrt{350} \le m_i(0) < \sqrt{500} \to \{\sqrt{100} \le m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	0.06/0.07/95.2	0.38/0.38/94.2	0.65/0.66/95	0.52/0.53/94.8	0.42/0.43/95.2	0.06/0.07/95.6	0.38/0.38/94	0.65/0.65/94.8	0.52/0.51/93.8	0.42/0.41/94.6
$\sqrt{350} \le m_i(0) < \sqrt{500} \to \{\sqrt{200} \le m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	0.25/0.26/95.2	0.76/0.76/94.2	1/1.01/94.6	0.83/0.84/94.8	0.71/0.71/95.4	0.25/0.26/95.2	0.76/0.76/94.2	1/0.99/94	0.83/0.82/94.4	0.71/0.69/94.6
$\sqrt{350} \le m_i(0) < \sqrt{500} \to \{\sqrt{250} \le m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	2.69/2.7/93.6	4.19/4.17/94.4	4.39/4.38/95	3.76/3.75/94.4	3.28/3.28/95.2	2.69/2.7/94	4.19/4.17/94.6	4.39/4.34/94.6	3.76/3.71/94	3.28/3.23/94.4
$\sqrt{350} \le m_i(0) < \sqrt{500} \to \{\sqrt{350} \le m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	17.2/17.17/94.2	15.37/15.27/95.2	13.16/13.06/94.2	11.6/11.51/93.2	10.39/10.32/93.8	17.2/17.21/94.4	15.37/15.29/95.2	13.16/13.03/93.6	11.6/11.46/92.4	10.39/10.26/93
$\sqrt{350} \le m_i(0) < \sqrt{500} \to \{\sqrt{500} \le m_i(t)\} \cap T_i^\star > t$	65.81/65.81/94	54.5/54.66/95.6	47.27/47.46/94.6	42.92/43.07/95	39.5/39.59/96.2	65.81/66.03/94.6	54.5/54.83/94.6	47.27/47.56/95.4	42.92/43.11/95	39.5/39.58/96.2
$\sqrt{500} \le m_i(0) \to \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	<0.01/<0.01/95	<0.01/<0.01/93.4	<0.01/<0.01/95	<0.01/<0.01/94.2	<0.01/<0.01/93.6	<0.01/<0.01/95.2	<0.01/<0.01/93.8	<0.01/<0.01/94.8	<0.01/<0.01/92.6	<0.01/<0.01/92.8
$\sqrt{500} \le m_i(0) \to \{\sqrt{50} \le m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	<0.01/<0.01/95.2	<0.01/<0.01/94	0.02/0.02/94.8	0.01/0.01/94	0.01/0.01/94.6	<0.01/<0.01/95.4	<0.01/<0.01/94.6	0.02/0.02/94	0.01/0.01/93	0.01/0.01/93.4
$\sqrt{500} \le m_i(0) \to \{\sqrt{100} \le m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	<0.01/<0.01/94.4	0.11/0.11/93.8	0.3/0.31/94.2	0.24/0.25/94.2	0.2/0.21/94.6	<0.01/<0.01/94.6	0.11/0.11/94.4	0.3/0.3/94.2	0.24/0.24/94.2	0.2/0.19/93.8
$\sqrt{500} \le m_i(0) \to \{\sqrt{200} \le m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	0.02/0.02/93.8	0.26/0.26/94.2	0.53/0.54/94.4	0.45/0.45/94.2	0.38/0.39/94.4	0.02/0.02/94.2	0.26/0.26/94.4	0.53/0.53/94.2	0.45/0.44/93.4	0.38/0.37/94
$\sqrt{500} \le m_i(0) \to \{\sqrt{250} \le m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	0.4/0.41/93.2	1.83/1.82/93.8	2.7/2.69/94.6	2.33/2.32/94.2	2.04/2.04/94.2	0.4/0.41/94.2	1.83/1.82/94	2.7/2.67/94.2	2.33/2.28/93.8	2.04/2/94.2
$\sqrt{500} \le m_i(0) \to \{\sqrt{350} \le m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	5.29/5.3/93.8	9.33/9.23/94.8	9.92/9.8/95	8.77/8.67/94.4	7.88/7.79/94.4	5.29/5.31/94	9.33/9.24/95	9.92/9.77/94	8.77/8.61/94	7.88/7.72/93.8
$\sqrt{500} \le m_i(0) \to \{\sqrt{500} \le m_i(t)\} \cap T_i^\star > t$	81.92/81.88/94.8	66.09/66.21/95.6	55.83/56.01/94.4	50.88/51.03/95.2	46.97/47.07/95.4	81.92/82.19/93.8	66.09/66.47/94.6	55.83/56.13/94.4	50.88/51.03/95	46.97/46.98/95.8

† Results from 500 replications with each dataset including 2000 individuals. The true marker evolution was based on linear splines with knots at 1 and 5 years since baseline and it was correctly specified in the fitted SREM-CIF-1 and SREM-CIF-2 models. The true parameter values, the mean of posterior medians over the 500 replications and the empirical coverage probability of posterior credible intervals are presented, respectively.

Table 5.19: Simulation study results for transition marker state probabilities (%) by baseline marker state for group 1 ($w = 1$) when the data have been generated by SREM-CIF-2 and there is misclassification†.

transition	$t=2$	$t=4$	$t=6$	$t=8$	$t=10$	$t=2$	$t=4$	$t=6$	$t=8$	$t=10$
	Results from SREM-CIF-1					Results from SREM-CIF-2				
$m_i(0) < \sqrt{50} \to \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	1.43/1.46/94.8	0.23/0.23/95.4	0.08/0.08/95.2	0.03/0.02/93.6	<0.01/0.01/92.8	1.43/1.45/95	0.23/0.24/96	0.08/0.09/94.8	0.03/0.03/93.6	<0.01/0.01/94
$m_i(0) < \sqrt{50} \to \{\sqrt{50} \le m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	5.95/6.08/95	1.26/1.29/95.4	0.53/0.53/94.8	0.3/0.29/93.6	0.15/0.14/93.4	5.95/5.98/95.2	1.26/1.31/95.4	0.53/0.57/94.6	0.3/0.33/94	0.15/0.18/93.4
$m_i(0) < \sqrt{50} \to \{\sqrt{100} \le m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	24.39/24.69/92.8	8.11/8.29/95.2	3.94/4/94.8	2.9/2.88/93.4	2.14/2.09/93.2	24.39/24.36/93.8	8.11/8.24/96.2	3.94/4.06/95.2	2.9/2.99/95	2.14/2.22/94.8
$m_i(0) < \sqrt{50} \to \{\sqrt{200} \le m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	12.33/12.39/93.8	6.5/6.59/94	3.61/3.65/94.8	2.88/2.88/94	2.34/2.32/93.4	12.33/12.26/94.4	6.5/6.52/95	3.61/3.65/95.4	2.88/2.91/95.8	2.34/2.37/94.4
$m_i(0) < \sqrt{50} \to \{\sqrt{250} \le m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	15.39/15.4/94.2	13.94/14/93.8	9.18/9.21/93.4	7.66/7.65/93.4	6.52/6.47/93.2	15.39/15.26/95.2	13.94/13.87/94.4	9.18/9.18/94.2	7.66/7.66/94.2	6.52/6.51/94.6
$m_i(0) < \sqrt{50} \to \{\sqrt{350} \le m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	6.17/6.16/93.6	14.49/14.43/95.2	13.05/12.97/93.8	11.33/11.23/93	10.01/9.89/92.6	6.17/6.11/94	14.49/14.31/94.8	13.05/12.9/94.2	11.33/11.19/94	10.01/9.88/93.8
$m_i(0) < \sqrt{50} \to \{\sqrt{500} \le m_i(t)\} \cap T_i^\star > t$	0.65/0.66/94.2	7.38/7.34/94.8	12.55/12.47/95	11.27/11.18/95.2	10.27/10.17/94.8	0.65/0.65/93.8	7.38/7.29/94.6	12.55/12.41/94	11.27/11.13/94.4	10.27/10.12/93.6
$\sqrt{50} \le m_i(0) < \sqrt{100} \to \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	0.11/0.11/93	0.07/0.08/95.6	0.05/0.06/94.4	0.02/0.02/92.8	0.01/0.01/92.4	0.11/0.11/94.8	0.07/0.07/95.4	0.05/0.06/94	0.02/0.03/93.2	0.01/0.01/92.6
$\sqrt{50} \le m_i(0) < \sqrt{100} \to \{\sqrt{50} \le m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	1.26/1.29/93.8	0.54/0.57/94.2	0.36/0.38/94	0.24/0.25/92.2	0.15/0.16/93.4	1.26/1.28/94.8	0.54/0.57/95.2	0.36/0.38/95.6	0.24/0.25/93	0.15/0.16/93.6
$\sqrt{50} \le m_i(0) < \sqrt{100} \to \{\sqrt{100} \le m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	13.66/13.83/93.8	5.18/5.35/92.6	3.12/3.26/91.4	2.43/2.53/92	1.92/2/92.4	13.66/13.73/94.6	5.18/5.28/94.6	3.12/3.22/94.6	2.43/2.5/94	1.92/1.98/93.4
$\sqrt{50} \le m_i(0) < \sqrt{100} \to \{\sqrt{200} \le m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	12.56/12.63/94	5.36/5.46/92.2	3.25/3.33/91.8	2.67/2.74/91	2.25/2.31/91.6	12.56/12.58/94.2	5.36/5.41/93.6	3.25/3.3/94.4	2.67/2.71/94.6	2.25/2.28/93.6
$\sqrt{50} \le m_i(0) < \sqrt{100} \to \{\sqrt{250} \le m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	25.07/25.1/94.8	14.47/14.59/92.4	9.39/9.52/92.6	8.01/8.12/93.8	6.97/7.06/92.2	25.07/25.02/93.8	14.47/14.5/93.6	9.39/9.45/93.4	8.01/8.05/95.6	6.97/7.01/93.8
$\sqrt{50} \le m_i(0) < \sqrt{100} \to \{\sqrt{350} \le m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	18.18/18.12/94.8	20.3/20.27/94.6	15.93/15.94/93.4	14.03/14.03/94.2	12.57/12.57/94.6	18.18/18.08/95	20.3/20.21/94.6	15.93/15.87/94.2	14.03/13.97/93.6	12.57/12.5/94
$\sqrt{50} \le m_i(0) < \sqrt{100} \to \{\sqrt{500} \le m_i(t)\} \cap T_i^\star > t$	3.75/3.75/93.6	15.29/15.22/96	20.19/20.1/95.2	18.3/18.21/94.8	16.82/16.73/94.4	3.75/3.74/93	15.29/15.21/95.2	20.19/20.08/95.2	18.3/18.18/95.2	16.82/16.69/95.4
$\sqrt{100} \le m_i(0) < \sqrt{200} \to \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	0.01/0.01/93.6	0.02/0.02/95.2	0.03/0.03/94	0.01/0.02/92.6	0.01/0.01/92.4	0.01/0.01/95	0.02/0.02/95.2	0.03/0.03/94.8	0.01/0.02/92.6	0.01/0.01/93
$\sqrt{100} \le m_i(0) < \sqrt{200} \to \{\sqrt{50} \le m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	0.22/0.23/93	0.22/0.24/94.8	0.21/0.23/93	0.15/0.16/91.2	0.1/0.11/92	0.22/0.23/94	0.22/0.23/94.8	0.21/0.22/94.4	0.15/0.15/93.2	0.1/0.11/93.2
$\sqrt{100} \le m_i(0) < \sqrt{200} \to \{\sqrt{100} \le m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	4.86/4.9/94.4	2.88/2.96/93.2	2.2/2.3/91.2	1.74/1.83/90.6	1.41/1.49/90	4.86/4.89/93.8	2.88/2.93/95	2.2/2.26/94.6	1.74/1.79/92.6	1.41/1.45/93
$\sqrt{100} \le m_i(0) < \sqrt{200} \to \{\sqrt{200} \le m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	7.03/7.05/93	3.64/3.69/92.2	2.56/2.63/91.6	2.13/2.2/91.2	1.82/1.88/90.2	7.03/7.05/92.8	3.64/3.67/93.8	2.56/2.6/93.4	2.13/2.16/93.6	1.82/1.85/93.2
$\sqrt{100} \le m_i(0) < \sqrt{200} \to \{\sqrt{250} \le m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	22.35/22.34/93.6	12.05/12.12/93.6	8.34/8.44/93.2	7.18/7.28/91.6	6.31/6.42/92.4	22.35/22.36/94.4	12.05/12.08/93.4	8.34/8.38/94.2	7.18/7.21/94	6.31/6.34/94
$\sqrt{100} \le m_i(0) < \sqrt{200} \to \{\sqrt{350} \le m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	30.61/30.51/95	22.4/22.36/93.6	16.71/16.7/92.8	14.81/14.82/93.4	13.36/13.38/94.2	30.61/30.57/95.6	22.4/22.36/93	16.71/16.67/93	14.81/14.76/93.4	13.36/13.31/93.6
$\sqrt{100} \le m_i(0) < \sqrt{200} \to \{\sqrt{500} \le m_i(t)\} \cap T_i^\star > t$	14.21/14.17/95.4	25.69/25.61/95.4	28.15/28.07/95.8	25.66/25.58/95.2	23.71/23.64/95.4	14.21/14.19/95.6	25.69/25.67/95.2	28.15/28.12/95.6	25.66/25.6/95.8	23.71/23.63/95.2
$\sqrt{200} \le m_i(0) < \sqrt{250} \to \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	<0.01/<0.01/93.6	0.01/0.01/95	0.01/0.02/93.6	0.01/0.01/92.8	<0.01/0.01/92.6	<0.01/<0.01/94.2	0.01/0.01/95.6	0.01/0.02/96	0.01/0.01/94	<0.01/<0.01/92.4
$\sqrt{200} \le m_i(0) < \sqrt{250} \to \{\sqrt{50} \le m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	0.03/0.03/94	0.09/0.09/95.6	0.12/0.14/93.2	0.09/0.1/91.6	0.06/0.07/91.2	0.03/0.03/94.2	0.09/0.09/95.6	0.12/0.13/94.8	0.09/0.09/93.4	0.06/0.07/93
$\sqrt{200} \le m_i(0) < \sqrt{250} \to \{\sqrt{100} \le m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	1.29/1.31/94.2	1.52/1.56/94.6	1.51/1.58/92.8	1.21/1.28/91.8	1/1.06/90.4	1.29/1.31/93.8	1.52/1.55/95.6	1.51/1.55/93.8	1.21/1.24/92.8	1/1.02/92.8
$\sqrt{200} \le m_i(0) < \sqrt{250} \to \{\sqrt{200} \le m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	2.88/2.88/94	2.31/2.33/94.8	1.95/2/92.8	1.64/1.69/91.8	1.41/1.46/91.8	2.88/2.88/93.8	2.31/2.32/94	1.95/1.97/92.8	1.64/1.65/92.4	1.41/1.42/93
$\sqrt{200} \le m_i(0) < \sqrt{250} \to \{\sqrt{250} \le m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	14.34/14.27/94	9.19/9.2/93.2	7.07/7.12/92.2	6.11/6.19/91.4	5.4/5.49/93.2	14.34/14.32/93.4	9.19/9.2/93.2	7.07/7.08/92.4	6.11/6.12/93.4	5.4/5.41/93.4
$\sqrt{200} \le m_i(0) < \sqrt{250} \to \{\sqrt{350} \le m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	33.97/33.81/94	21.77/21.66/92.4	16.32/16.26/93.6	14.52/14.5/93.2	13.15/13.15/94.4	33.97/33.94/94	21.77/21.71/92.8	16.32/16.25/92.6	14.52/14.45/92.8	13.15/13.08/93.6
$\sqrt{200} \le m_i(0) < \sqrt{250} \to \{\sqrt{500} \le m_i(t)\} \cap T_i^\star > t$	29.69/29.6/95.6	35.72/35.61/95	35.13/35.04/95	32.12/32.06/95.4	29.76/29.73/95	29.69/29.71/95.8	35.72/35.76/95.4	35.13/35.17/95.2	32.12/32.12/95.2	29.76/29.74/95.6
$\sqrt{250} \le m_i(0) < \sqrt{350} \to \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	<0.01/<0.01/93.4	<0.01/<0.01/95	0.01/0.01/94.4	0.01/0.01/92.6	<0.01/<0.01/92	<0.01/<0.01/93.4	<0.01/<0.01/95.4	0.01/0.01/95.4	0.01/0.01/94.4	<0.01/<0.01/92.6
$\sqrt{250} \le m_i(0) < \sqrt{350} \to \{\sqrt{50} \le m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	0.01/0.01/93.6	0.04/0.04/94.8	0.08/0.09/94.6	0.06/0.06/93	0.04/0.05/92.2	0.01/0.01/93.6	0.04/0.04/95.4	0.08/0.08/95.2	0.06/0.06/93.8	0.04/0.04/93.2
$\sqrt{250} \le m_i(0) < \sqrt{350} \to \{\sqrt{100} \le m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	0.42/0.43/93.2	0.89/0.91/94.6	1.09/1.14/94.8	0.88/0.93/93.2	0.73/0.78/91.2	0.42/0.43/93.4	0.89/0.9/95.2	1.09/1.12/95.2	0.88/0.9/94.2	0.73/0.75/93.6
$\sqrt{250} \le m_i(0) < \sqrt{350} \to \{\sqrt{200} \le m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	1.2/1.2/93.4	1.52/1.53/95	1.52/1.55/94.4	1.28/1.32/94	1.11/1.14/91.8	1.2/1.2/93	1.52/1.53/94.2	1.52/1.53/93.6	1.28/1.29/94.2	1.11/1.11/93
$\sqrt{250} \le m_i(0) < \sqrt{350} \to \{\sqrt{250} \le m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	8.1/8.05/94	6.91/6.9/93.4	5.95/5.98/92.8	5.17/5.21/92.8	4.58/4.64/92.2	8.1/8.08/94.2	6.91/6.9/92.6	5.95/5.95/92.4	5.17/5.15/93.6	4.58/4.57/93.8
$\sqrt{250} \le m_i(0) < \sqrt{350} \to \{\sqrt{350} \le m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	28.95/28.72/94.6	19.65/19.48/93.4	15.33/15.23/93.4	13.67/13.61/93.6	12.4/12.38/94	28.95/28.86/95.4	19.65/19.54/94.2	15.33/15.23/93.2	13.67/13.56/93.4	12.4/12.3/93.8
$\sqrt{250} \le m_i(0) < \sqrt{350} \to \{\sqrt{500} \le m_i(t)\} \cap T_i^\star > t$	45.21/45.04/95.2	43.83/43.68/94.8	40.55/40.45/94.2	37.15/37.09/95.6	34.5/34.47/96	45.21/45.27/94.6	43.83/43.91/94.6	40.55/40.62/94.8	37.15/37.18/96	34.5/34.48/95.8
$\sqrt{350} \le m_i(0) < \sqrt{500} \to \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	<0.01/<0.01/93.6	<0.01/<0.01/94.8	<0.01/<0.01/94.8	<0.01/<0.01/93.8	<0.01/<0.01/92.8	<0.01/<0.01/93.4	<0.01/<0.01/95.4	<0.01/<0.01/95.8	<0.01/<0.01/94.6	<0.01/<0.01/93.6
$\sqrt{350} \le m_i(0) < \sqrt{500} \to \{\sqrt{50} \le m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	<0.01/<0.01/93.6	0.01/0.01/94.4	0.04/0.04/95.4	0.03/0.03/94	0.02/0.03/93	<0.01/<0.01/93.6	0.01/0.01/94.8	0.04/0.04/95.8	0.03/0.03/96.2	0.02/0.02/94.4
$\sqrt{350} \le m_i(0) < \sqrt{500} \to \{\sqrt{100} \le m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	0.06/0.06/93.6	0.37/0.38/95	0.65/0.68/94.4	0.53/0.56/94.6	0.44/0.47/93.8	0.06/0.06/93.6	0.37/0.38/94.8	0.65/0.67/95.2	0.53/0.54/94.8	0.44/0.45/94.6
$\sqrt{350} \le m_i(0) < \sqrt{500} \to \{\sqrt{200} \le m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	0.25/0.25/93	0.76/0.76/95	1.01/1.02/94.6	0.85/0.87/94.4	0.74/0.76/93.6	0.25/0.25/93.4	0.76/0.76/95.4	1.01/1.01/94.6	0.85/0.85/94.4	0.74/0.74/94.2
$\sqrt{350} \le m_i(0) < \sqrt{500} \to \{\sqrt{250} \le m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	2.67/2.65/94	4.18/4.15/95	4.42/4.42/94.2	3.85/3.87/94	3.42/3.46/93.2	2.67/2.66/93.6	4.18/4.15/94.4	4.42/4.4/93.6	3.85/3.82/93.4	3.42/3.4/93.4
$\sqrt{350} \le m_i(0) < \sqrt{500} \to \{\sqrt{350} \le m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	17.12/16.92/95.6	15.38/15.16/93.2	13.31/13.15/92.6	11.89/11.79/92.4	10.81/10.75/93.2	17.12/17/95.4	15.38/15.22/92.8	13.31/13.15/92	11.89/11.73/91.4	10.81/10.66/92.6
$\sqrt{350} \le m_i(0) < \sqrt{500} \to \{\sqrt{500} \le m_i(t)\} \cap T_i^\star > t$	65.65/65.29/94.2	54.69/54.44/94	47.91/47.77/93.8	44/43.92/94.2	40.96/40.93/94.4	65.65/65.67/94.6	54.69/54.76/94	47.91/47.99/93.4	44/44.03/94.6	40.96/40.92/95
$\sqrt{500} \le m_i(0) \to \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	<0.01/<0.01/93.8	<0.01/<0.01/95	<0.01/<0.01/95.6	<0.01/<0.01/95.2	<0.01/<0.01/93.8	<0.01/<0.01/93.8	<0.01/<0.01/94.6	<0.01/<0.01/95.8	<0.01/<0.01/95	<0.01/<0.01/93.2
$\sqrt{500} \le m_i(0) \to \{\sqrt{50} \le m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	<0.01/<0.01/94	<0.01/<0.01/95	0.02/0.02/95.6	0.01/0.01/94.4	0.01/0.01/93.8	<0.01/<0.01/93.6	<0.01/<0.01/94.8	0.02/0.02/94.8	0.01/0.01/95.6	0.01/0.01/93.8
$\sqrt{500} \le m_i(0) \to \{\sqrt{100} \le m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	<0.01/<0.01/93.2	0.1/0.11/95	0.31/0.32/94.4	0.25/0.26/94.4	0.21/0.23/93.6	<0.01/<0.01/93.4	0.1/0.11/95	0.31/0.31/94.8	0.25/0.25/95.4	0.21/0.21/94
$\sqrt{500} \le m_i(0) \to \{\sqrt{200} \le m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	0.02/0.02/93.4	0.26/0.26/94.6	0.54/0.55/94.4	0.46/0.47/95	0.4/0.41/94	0.02/0.02/93.2	0.26/0.26/95	0.54/0.54/94.6	0.46/0.46/94	0.4/0.4/93.6
$\sqrt{500} \le m_i(0) \to \{\sqrt{250} \le m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	0.39/0.4/93.6	1.83/1.81/94.2	2.73/2.71/94.8	2.38/2.38/94	2.12/2.14/93.6	0.39/0.4/93.2	1.83/1.81/94.2	2.73/2.7/94.6	2.38/2.35/93	2.12/2.1/93.2
$\sqrt{500} \le m_i(0) \to \{\sqrt{350} \le m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	5.27/5.21/94.6	9.35/9.15/93.4	10.04/9.86/93	8.99/8.86/93.8	8.19/8.1/93.2	5.27/5.23/94.6	9.35/9.18/94.2	10.04/9.85/93.2	8.99/8.8/92.6	8.19/8.01/93.2
$\sqrt{500} \le m_i(0) \to \{\sqrt{500} \le m_i(t)\} \cap T_i^\star > t$	81.83/81.19/92	66.4/65.93/93	56.6/56.35/93.8	52.13/51.99/94.6	48.63/48.58/95	81.83/81.67/93.8	66.4/66.34/93.6	56.6/56.6/93.8	52.13/52.08/93.8	48.63/48.53/95

† Results from 500 replications with each dataset including 2000 individuals. The true marker evolution was based on linear splines with knots at 1 and 5 years since baseline and it was correctly specified in the fitted SREM-CIF-1 and SREM-CIF-2 models. The true parameter values, the mean of posterior medians over the 500 replications and the empirical coverage probability of posterior credible intervals are presented, respectively.

Table 5.20: Simulation study results for transition marker state probabilities (%) by baseline marker state for group 0 ($w = 0$) when the data have been generated by SREM-CIF-2 and there is misclassification†.

transition	$t=2$	$t=4$	$t=6$	$t=8$	$t=10$	$t=2$	$t=4$	$t=6$	$t=8$	$t=10$
	Results from SREM-CIF-1					Results from SREM-CIF-2				
$m_i(0) < \sqrt{50} \to \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	1.47/1.52/94.8	0.23/0.24/95.4	0.08/0.09/95	0.03/0.03/93.8	<0.01/0.01/93	1.47/1.49/95.4	0.23/0.25/96	0.08/0.1/94.4	0.03/0.03/93.6	<0.01/0.01/94
$m_i(0) < \sqrt{50} \to \{\sqrt{50} \le m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	6.04/6.22/94.2	1.28/1.34/94.4	0.54/0.56/93.8	0.31/0.31/94.2	0.15/0.15/93.2	6.04/6.09/95.2	1.28/1.34/95.4	0.54/0.58/94	0.31/0.33/93.2	0.15/0.18/94.2
$m_i(0) < \sqrt{50} \to \{\sqrt{100} \le m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	24.53/24.94/91	8.15/8.42/92.6	3.94/4.07/92.4	2.89/2.93/93	2.12/2.12/94	24.53/24.56/92.2	8.15/8.3/94.4	3.94/4.08/92.8	2.89/2.99/93.4	2.12/2.22/93.4
$m_i(0) < \sqrt{50} \to \{\sqrt{200} \le m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	12.34/12.44/92.4	6.5/6.62/91.4	3.59/3.66/91	2.85/2.88/92.6	2.3/2.3/93	12.34/12.28/93.6	6.5/6.53/93.6	3.59/3.64/93.6	2.85/2.89/93.2	2.3/2.34/93.6
$m_i(0) < \sqrt{50} \to \{\sqrt{250} \le m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	15.36/15.41/95.2	13.87/14/92.8	9.09/9.18/91.6	7.54/7.56/92.2	6.38/6.35/93.4	15.36/15.24/95.4	13.87/13.82/93.8	9.09/9.11/92	7.54/7.56/92.6	6.38/6.4/93.4
$m_i(0) < \sqrt{50} \to \{\sqrt{350} \le m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	6.14/6.15/94.6	14.34/14.34/94.4	12.84/12.82/93.6	11.08/11.01/94.6	9.74/9.62/94	6.14/6.08/94.4	14.34/14.19/94.2	12.84/12.72/93.2	11.08/10.97/94	9.74/9.64/94.2
$m_i(0) < \sqrt{50} \to \{\sqrt{500} \le m_i(t)\} \cap T_i^\star > t$	0.64/0.65/94	7.27/7.26/94.8	12.29/12.25/95.2	10.96/10.88/94.6	9.94/9.82/95.8	0.64/0.65/93.8	7.27/7.19/94.8	12.29/12.16/94.6	10.96/10.84/95.2	9.94/9.8/95.2
$\sqrt{50} \le m_i(0) < \sqrt{100} \to \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	0.11/0.12/94.4	0.07/0.08/95	0.05/0.06/94.4	0.02/0.03/93.2	0.01/0.01/92.2	0.11/0.11/94.8	0.07/0.08/96.4	0.05/0.06/94.6	0.02/0.03/93.6	0.01/0.01/93.6
$\sqrt{50} \le m_i(0) < \sqrt{100} \to \{\sqrt{50} \le m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	1.26/1.31/94.4	0.54/0.58/94.6	0.36/0.39/92.6	0.24/0.25/93.2	0.15/0.16/93.6	1.26/1.29/94.8	0.54/0.57/96	0.36/0.38/94.8	0.24/0.25/94.2	0.15/0.16/93.4
$\sqrt{50} \le m_i(0) < \sqrt{100} \to \{\sqrt{100} \le m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	13.65/13.86/94.4	5.16/5.35/93.2	3.09/3.25/91.6	2.39/2.51/92.6	1.89/1.97/92.2	13.65/13.75/94.6	5.16/5.27/95.2	3.09/3.2/93.6	2.39/2.48/93.8	1.89/1.96/93.4
$\sqrt{50} \le m_i(0) < \sqrt{100} \to \{\sqrt{200} \le m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	12.51/12.62/92.2	5.31/5.43/90.6	3.2/3.3/91.2	2.62/2.7/92	2.19/2.25/92.4	12.51/12.54/92.6	5.31/5.37/94.8	3.2/3.26/92.6	2.62/2.67/92.6	2.19/2.24/93.6
$\sqrt{50} \le m_i(0) < \sqrt{100} \to \{\sqrt{250} \le m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	24.9/25.01/93	14.29/14.45/91.6	9.22/9.37/92.6	7.82/7.93/92.4	6.77/6.85/93	24.9/24.88/92.8	14.29/14.34/94	9.22/9.29/92.6	7.82/7.87/93	6.77/6.82/93.6
$\sqrt{50} \le m_i(0) < \sqrt{100} \to \{\sqrt{350} \le m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	18.02/18.02/94	19.97/20.01/93.4	15.57/15.61/94.2	13.63/13.62/95.2	12.14/12.11/94.6	18.02/17.93/94.4	19.97/19.9/93.8	15.57/15.53/95	13.63/13.58/95.2	12.14/12.1/94.8
$\sqrt{50} \le m_i(0) < \sqrt{100} \to \{\sqrt{500} \le m_i(t)\} \cap T_i^\star > t$	3.71/3.72/94	14.99/14.97/95.6	19.65/19.6/95.8	17.69/17.61/95.8	16.17/16.05/96.2	3.71/3.7/93.4	14.99/14.92/95	19.65/19.56/95.4	17.69/17.6/96.2	16.17/16.06/95.6
$\sqrt{100} \le m_i(0) < \sqrt{200} \to \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	0.01/0.01/94.8	0.02/0.03/95.4	0.03/0.03/93.6	0.01/0.02/93.2	0.01/0.01/93.2	0.01/0.01/94.8	0.02/0.02/96	0.03/0.03/94.4	0.01/0.02/93.6	0.01/0.01/94.2
$\sqrt{100} \le m_i(0) < \sqrt{200} \to \{\sqrt{50} \le m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	0.22/0.23/93.8	0.22/0.24/95	0.21/0.23/92.8	0.15/0.16/93.2	0.1/0.11/92.8	0.22/0.23/95	0.22/0.23/95.6	0.21/0.22/94.4	0.15/0.15/94	0.1/0.11/93.4
$\sqrt{100} \le m_i(0) < \sqrt{200} \to \{\sqrt{100} \le m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	4.84/4.89/94.6	2.85/2.94/94.8	2.16/2.27/92.6	1.71/1.8/92.2	1.38/1.46/91.6	4.84/4.87/95	2.85/2.91/95.6	2.16/2.23/94.8	1.71/1.76/94.8	1.38/1.43/93.8
$\sqrt{100} \le m_i(0) < \sqrt{200} \to \{\sqrt{200} \le m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	6.98/7.02/94	3.59/3.65/94.2	2.51/2.58/92.4	2.08/2.14/91.6	1.77/1.82/91.4	6.98/7/94	3.59/3.62/95.2	2.51/2.55/94	2.08/2.11/93.8	1.77/1.8/93.2
$\sqrt{100} \le m_i(0) < \sqrt{200} \to \{\sqrt{250} \le m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	22.15/22.19/93.6	11.85/11.94/93.8	8.15/8.26/92.4	6.97/7.07/93	6.1/6.18/93.8	22.15/22.17/93	11.85/11.89/94.2	8.15/8.2/93.4	6.97/7.01/94.2	6.1/6.14/94.8
$\sqrt{100} \le m_i(0) < \sqrt{200} \to \{\sqrt{350} \le m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	30.27/30.26/93.8	21.96/21.97/94	16.26/16.28/94	14.32/14.32/94.2	12.85/12.83/93.4	30.27/30.24/92	21.96/21.93/94.4	16.26/16.23/94.4	14.32/14.29/94.6	12.85/12.82/93.2
$\sqrt{100} \le m_i(0) < \sqrt{200} \to \{\sqrt{500} \le m_i(t)\} \cap T_i^\star > t$	14.03/14.03/95.6	25.11/25.1/94	27.3/27.27/94.4	24.71/24.65/95.2	22.71/22.61/95	14.03/14.02/95.2	25.11/25.1/94.4	27.3/27.28/93.8	24.71/24.67/94.2	22.71/22.65/94.6
$\sqrt{200} \le m_i(0) < \sqrt{250} \to \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	<0.01/<0.01/94.4	0.01/0.01/96.4	0.01/0.02/93	0.01/0.01/93.6	<0.01/0.01/92.8	<0.01/<0.01/94.4	0.01/0.01/96	0.01/0.02/94.4	0.01/0.01/94.6	<0.01/<0.01/94
$\sqrt{200} \le m_i(0) < \sqrt{250} \to \{\sqrt{50} \le m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	0.03/0.03/94.8	0.09/0.09/96.2	0.12/0.14/93.8	0.09/0.1/93	0.06/0.07/93.4	0.03/0.03/95.2	0.09/0.09/96	0.12/0.13/95	0.09/0.09/94.8	0.06/0.07/94.6
$\sqrt{200} \le m_i(0) < \sqrt{250} \to \{\sqrt{100} \le m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	1.29/1.3/95	1.5/1.55/96.4	1.48/1.55/94	1.19/1.25/93.4	0.97/1.03/92	1.29/1.3/94.6	1.5/1.53/96.4	1.48/1.52/95.8	1.19/1.21/94.8	0.97/1/94.4
$\sqrt{200} \le m_i(0) < \sqrt{250} \to \{\sqrt{200} \le m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	2.85/2.86/95.4	2.27/2.3/97	1.91/1.95/94.8	1.59/1.64/94.8	1.36/1.41/94	2.85/2.86/95	2.27/2.28/96.4	1.91/1.93/95.6	1.59/1.61/95.2	1.36/1.38/95
$\sqrt{200} \le m_i(0) < \sqrt{250} \to \{\sqrt{250} \le m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	14.18/14.16/96.2	9.01/9.04/95.2	6.88/6.94/94.8	5.92/5.98/95	5.21/5.27/95.8	14.18/14.17/95.2	9.01/9.02/96.2	6.88/6.9/95.6	5.92/5.93/95	5.21/5.22/95
$\sqrt{200} \le m_i(0) < \sqrt{250} \to \{\sqrt{350} \le m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	33.55/33.48/93.2	21.29/21.24/92.8	15.84/15.81/94	14/13.97/93.6	12.61/12.58/92.8	33.55/33.53/93.6	21.29/21.24/92.6	15.84/15.78/93.6	14/13.94/93.6	12.61/12.55/94
$\sqrt{200} \le m_i(0) < \sqrt{250} \to \{\sqrt{500} \le m_i(t)\} \cap T_i^\star > t$	29.28/29.27/95.8	34.86/34.85/93.6	34/33.98/93.8	30.87/30.83/94.6	28.45/28.38/94.6	29.28/29.31/95.6	34.86/34.9/93	34/34.04/93.4	30.87/30.88/93.6	28.45/28.43/94
$\sqrt{250} \le m_i(0) < \sqrt{350} \to \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	<0.01/<0.01/93.4	<0.01/<0.01/96	0.01/0.01/94.8	0.01/0.01/94.2	<0.01/<0.01/93.4	<0.01/<0.01/94	<0.01/<0.01/95.8	0.01/0.01/95.2	0.01/0.01/94.8	<0.01/<0.01/94.6
$\sqrt{250} \le m_i(0) < \sqrt{350} \to \{\sqrt{50} \le m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	0.01/0.01/93.8	0.04/0.04/96	0.08/0.09/94.2	0.06/0.06/93	0.04/0.05/92.8	0.01/0.01/94.2	0.04/0.04/95.8	0.08/0.08/96	0.06/0.06/94.4	0.04/0.04/94
$\sqrt{250} \le m_i(0) < \sqrt{350} \to \{\sqrt{100} \le m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	0.42/0.42/94	0.88/0.9/95.4	1.07/1.12/95	0.86/0.91/93	0.71/0.75/92.2	0.42/0.42/93.6	0.88/0.89/95.6	1.07/1.09/95.2	0.86/0.88/94.4	0.71/0.73/94.8
$\sqrt{250} \le m_i(0) < \sqrt{350} \to \{\sqrt{200} \le m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	1.18/1.19/93.8	1.49/1.51/95	1.48/1.51/95	1.24/1.27/94.4	1.07/1.1/93.8	1.18/1.19/93	1.49/1.5/95.2	1.48/1.5/95.2	1.24/1.25/95	1.07/1.08/95.2
$\sqrt{250} \le m_i(0) < \sqrt{350} \to \{\sqrt{250} \le m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	8.01/7.98/94.8	6.77/6.77/94.8	5.79/5.82/94.6	4.99/5.03/94.4	4.41/4.45/95.2	8.01/7.99/94.2	6.77/6.76/95	5.79/5.78/94.6	4.99/4.99/94.6	4.41/4.4/94.2
$\sqrt{250} \le m_i(0) < \sqrt{350} \to \{\sqrt{350} \le m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	28.57/28.42/95.4	19.19/19.08/94.4	14.86/14.78/93.8	13.16/13.1/93.8	11.88/11.82/93	28.57/28.48/95.8	19.19/19.09/94.8	14.86/14.76/93.6	13.16/13.06/93.4	11.88/11.79/93.2
$\sqrt{250} \le m_i(0) < \sqrt{350} \to \{\sqrt{500} \le m_i(t)\} \cap T_i^\star > t$	44.56/44.52/94	42.73/42.71/93.2	39.2/39.19/94.6	35.67/35.64/93.8	32.93/32.88/94.2	44.56/44.63/93.6	42.73/42.8/93.8	39.2/39.27/94.8	35.67/35.7/94.2	32.93/32.92/94
$\sqrt{350} \le m_i(0) < \sqrt{500} \to \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	<0.01/<0.01/94	<0.01/<0.01/96.2	<0.01/<0.01/95	<0.01/<0.01/93.8	<0.01/<0.01/93.2	<0.01/<0.01/93.8	<0.01/<0.01/96	<0.01/<0.01/95.8	<0.01/<0.01/94.8	<0.01/<0.01/94.4
$\sqrt{350} \le m_i(0) < \sqrt{500} \to \{\sqrt{50} \le m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	<0.01/<0.01/94.2	0.01/0.01/95.6	0.04/0.04/94.8	0.03/0.03/93.8	0.02/0.02/93.2	<0.01/<0.01/93.8	0.01/0.01/95.2	0.04/0.04/96.2	0.03/0.03/95	0.02/0.02/94.4
$\sqrt{350} \le m_i(0) < \sqrt{500} \to \{\sqrt{100} \le m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	0.06/0.06/93.2	0.37/0.38/95	0.64/0.66/95.4	0.52/0.54/94.2	0.43/0.46/93.8	0.06/0.06/93.8	0.37/0.37/94.8	0.64/0.65/95.4	0.52/0.53/95.4	0.43/0.44/94.2
$\sqrt{350} \le m_i(0) < \sqrt{500} \to \{\sqrt{200} \le m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	0.25/0.25/93.2	0.74/0.75/94.8	0.98/1/95.6	0.83/0.85/95	0.71/0.73/94.4	0.25/0.25/93.4	0.74/0.74/94.6	0.98/0.99/95.6	0.83/0.83/94.6	0.71/0.72/95
$\sqrt{350} \le m_i(0) < \sqrt{500} \to \{\sqrt{250} \le m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	2.63/2.63/94.4	4.08/4.06/95	4.29/4.29/94.4	3.71/3.73/94.4	3.29/3.31/94.4	2.63/2.63/94.2	4.08/4.06/94.8	4.29/4.27/93.8	3.71/3.69/94.4	3.29/3.27/94.8
$\sqrt{350} \le m_i(0) < \sqrt{500} \to \{\sqrt{350} \le m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	16.88/16.72/95.2	15/14.83/94.4	12.88/12.75/94.4	11.43/11.33/94	10.34/10.25/92.8	16.88/16.76/95.6	15/14.84/93.6	12.88/12.72/93	11.43/11.28/94	10.34/10.2/93.4
$\sqrt{350} \le m_i(0) < \sqrt{500} \to \{\sqrt{500} \le m_i(t)\} \cap T_i^\star > t$	64.67/64.49/92.8	53.28/53.18/94	46.26/46.24/94.4	42.19/42.17/93.8	39.05/39.01/93.4	64.67/64.69/92	53.28/53.34/93.6	46.26/46.34/94.4	42.19/42.22/94.2	39.05/39.02/93.4
$\sqrt{500} \le m_i(0) \to \{m_i(t) < \sqrt{50}\} \cap T_i^\star > t$	<0.01/<0.01/94	<0.01/<0.01/95	<0.01/<0.01/95.8	<0.01/<0.01/94.2	<0.01/<0.01/94	<0.01/<0.01/94	<0.01/<0.01/94.8	<0.01/<0.01/95.8	<0.01/<0.01/95.6	<0.01/<0.01/94
$\sqrt{500} \le m_i(0) \to \{\sqrt{50} \le m_i(t) < \sqrt{100}\} \cap T_i^\star > t$	<0.01/<0.01/93.8	<0.01/<0.01/94.8	0.01/0.02/95	0.01/0.01/94.2	0.01/0.01/94.4	<0.01/<0.01/93.8	<0.01/<0.01/95	0.01/0.02/95.6	0.01/0.01/95.2	0.01/0.01/94.6
$\sqrt{500} \le m_i(0) \to \{\sqrt{100} \le m_i(t) < \sqrt{200}\} \cap T_i^\star > t$	<0.01/<0.01/93.2	0.1/0.11/94.8	0.3/0.31/94.8	0.24/0.26/94	0.2/0.22/94.6	<0.01/<0.01/93.6	0.1/0.11/94.8	0.3/0.3/95	0.24/0.25/93.6	0.2/0.21/93.6
$\sqrt{500} \le m_i(0) \to \{\sqrt{200} \le m_i(t) < \sqrt{250}\} \cap T_i^\star > t$	0.02/0.02/93.8	0.25/0.26/95	0.52/0.53/94.8	0.44/0.45/94.4	0.38/0.39/94.6	0.02/0.02/93.6	0.25/0.26/94.8	0.52/0.52/94.4	0.44/0.44/94.4	0.38/0.38/94.4
$\sqrt{500} \le m_i(0) \to \{\sqrt{250} \le m_i(t) < \sqrt{350}\} \cap T_i^\star > t$	0.39/0.4/93.8	1.78/1.77/94.2	2.64/2.63/94.4	2.29/2.29/94.6	2.03/2.04/94.6	0.39/0.4/93.4	1.78/1.77/94.2	2.64/2.61/94	2.29/2.26/94.2	2.03/2.01/94.6
$\sqrt{500} \le m_i(0) \to \{\sqrt{350} \le m_i(t) < \sqrt{500}\} \cap T_i^\star > t$	5.19/5.15/94.6	9.11/8.94/94.4	9.7/9.54/93.4	8.63/8.5/94	7.82/7.71/94	5.19/5.15/94.6	9.11/8.95/94.4	9.7/9.52/93.2	8.63/8.45/93.2	7.82/7.65/92.6
$\sqrt{500} \le m_i(0) \to \{\sqrt{500} \le m_i(t)\} \cap T_i^\star > t$	80.59/80.16/91.4	64.65/64.38/93.6	54.62/54.52/94.2	49.95/49.89/94.4	46.34/46.28/93	80.59/80.44/91.4	64.65/64.59/93.4	54.62/54.61/93.6	49.95/49.9/94.2	46.34/46.22/93.6

† Results from 500 replications with each dataset including 2000 individuals. The true marker evolution was based on linear splines with knots at 1 and 5 years since baseline and it was correctly specified in the fitted SREM-CIF-1 and SREM-CIF-2 models. The true parameter values, the mean of posterior medians over the 500 replications and the empirical coverage probability of posterior credible intervals are presented, respectively.

5.6.2.3 Simulation study results for population-averaged CIFs by baseline marker state under misclassified failure cause

Results for population-averaged CIFs by baseline marker state are provided in Figures 5.5-5.8. Both models, even when the fitted model was misspecified, yielded almost unbiased estimates with satisfactory coverage rates. However, there were few cases where the coverage probabilities from the misspecified model were lower than the nominal level. For example, when data were generated from the SREM-CIF-1 model, the coverage probability based on the SREM-CIF-2 model of the population-averaged CIF of cause 1, conditional on the $m_i(0) > \sqrt{500}$ initial state, reduced to 87.6% (Figure 5.6). This result is very similar to that found in the simulation study assuming that the failure cause is available for all individuals (Figure 5.2).

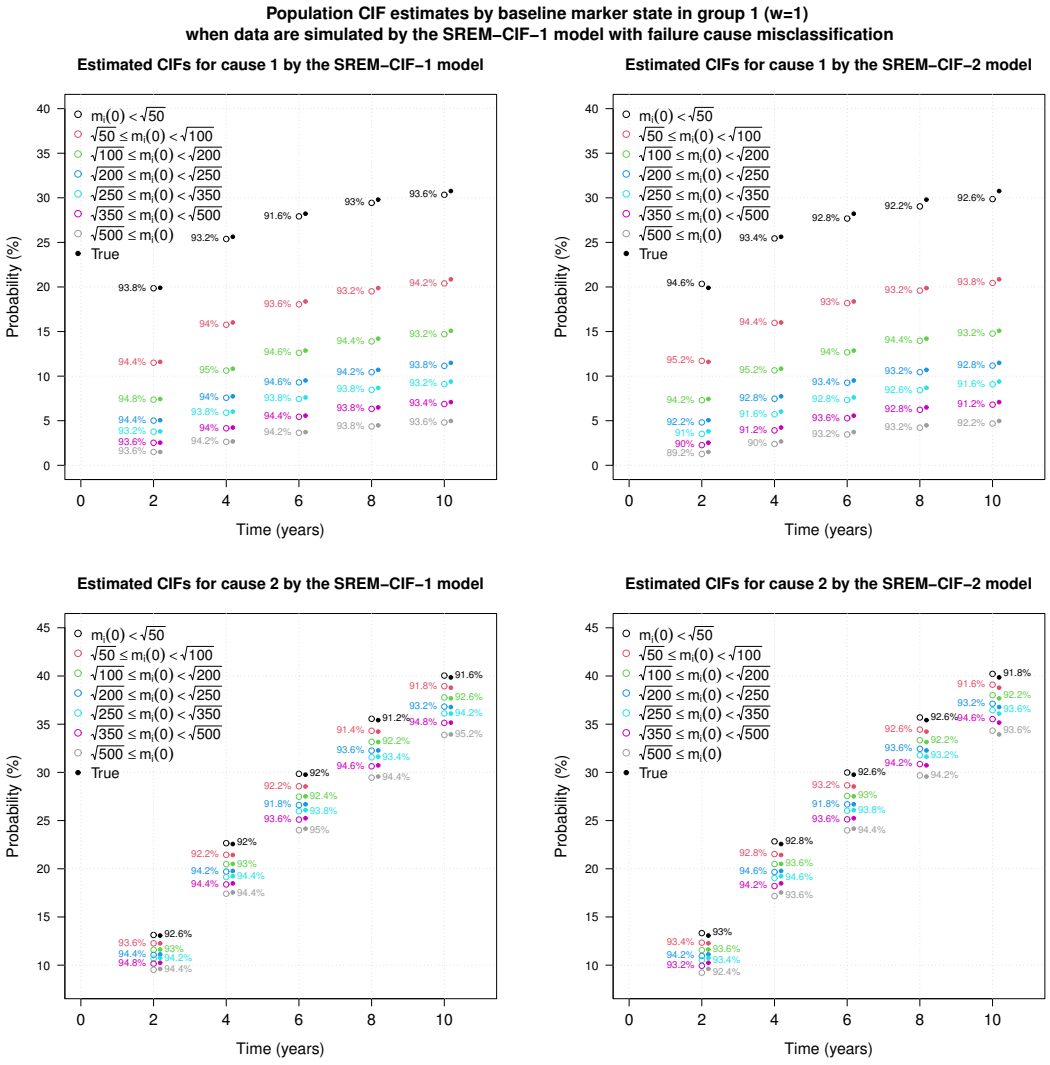


Figure 5.5: Simulation study results for population-averaged CIF estimates by baseline marker state for group 1 when data are simulated under the SREM-CIF-1 model and there is failure cause misclassification. CIF is estimated at certain years since baseline. Open circles show the empirical estimates based on posterior medians over 500 replications whereas closed circles show the true values. Shown are also the corresponding empirical coverage probabilities. The true marker is based on linear splines with knots at 1 and 5 years since baseline and it is correctly specified when fitting SREM-CIF-1 and SREM-CIF-2 models.

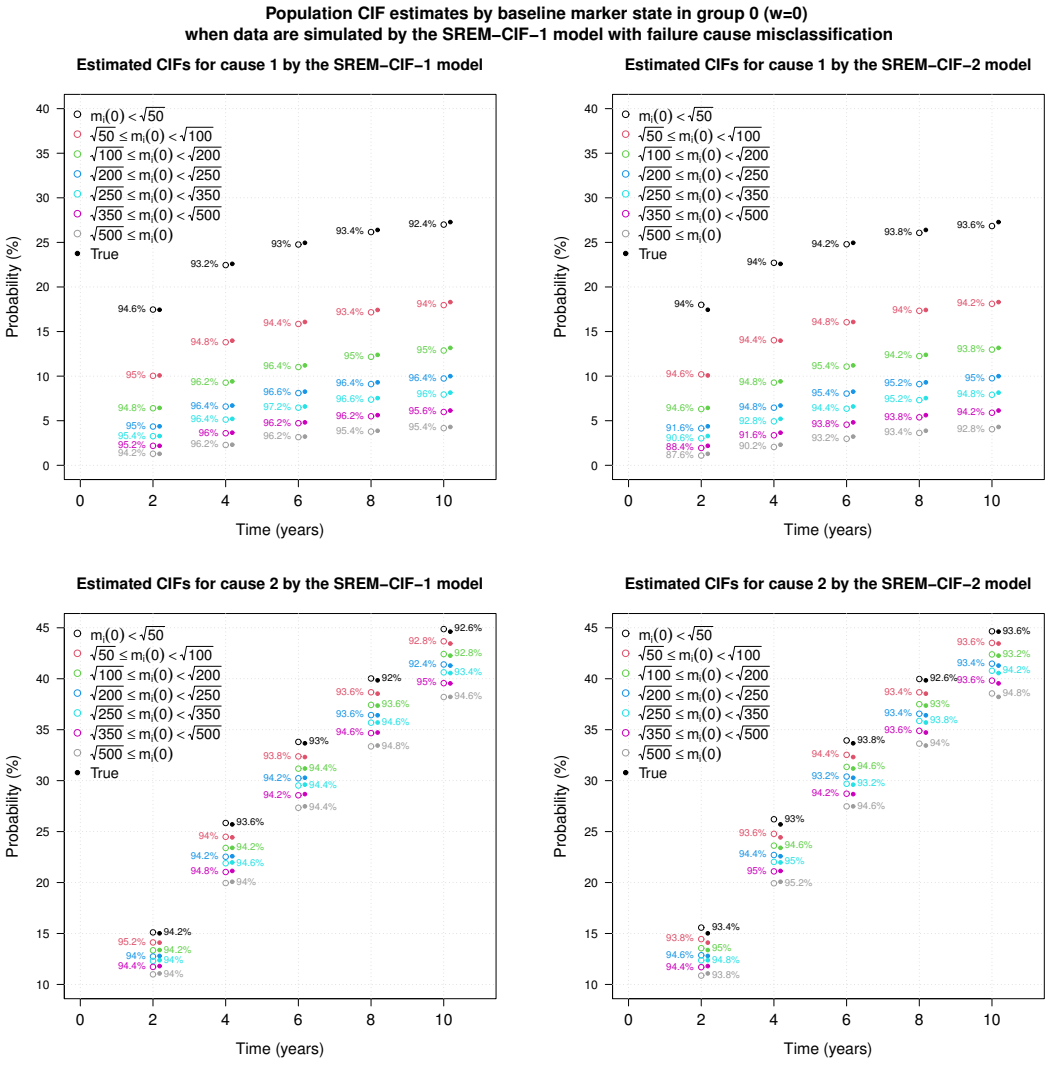


Figure 5.6: Simulation study results for population-averaged CIF estimates by baseline marker state for group 0 when data are simulated under the SREM-CIF-1 model and there is failure cause misclassification. CIF is estimated at certain years since baseline. Open circles show the empirical estimates based on posterior medians over 500 replications whereas closed circles show the true values. Shown are also the corresponding empirical coverage probabilities. The true marker is based on linear splines with knots at 1 and 5 years since baseline and it is correctly specified when fitting SREM-CIF-1 and SREM-CIF-2 models.

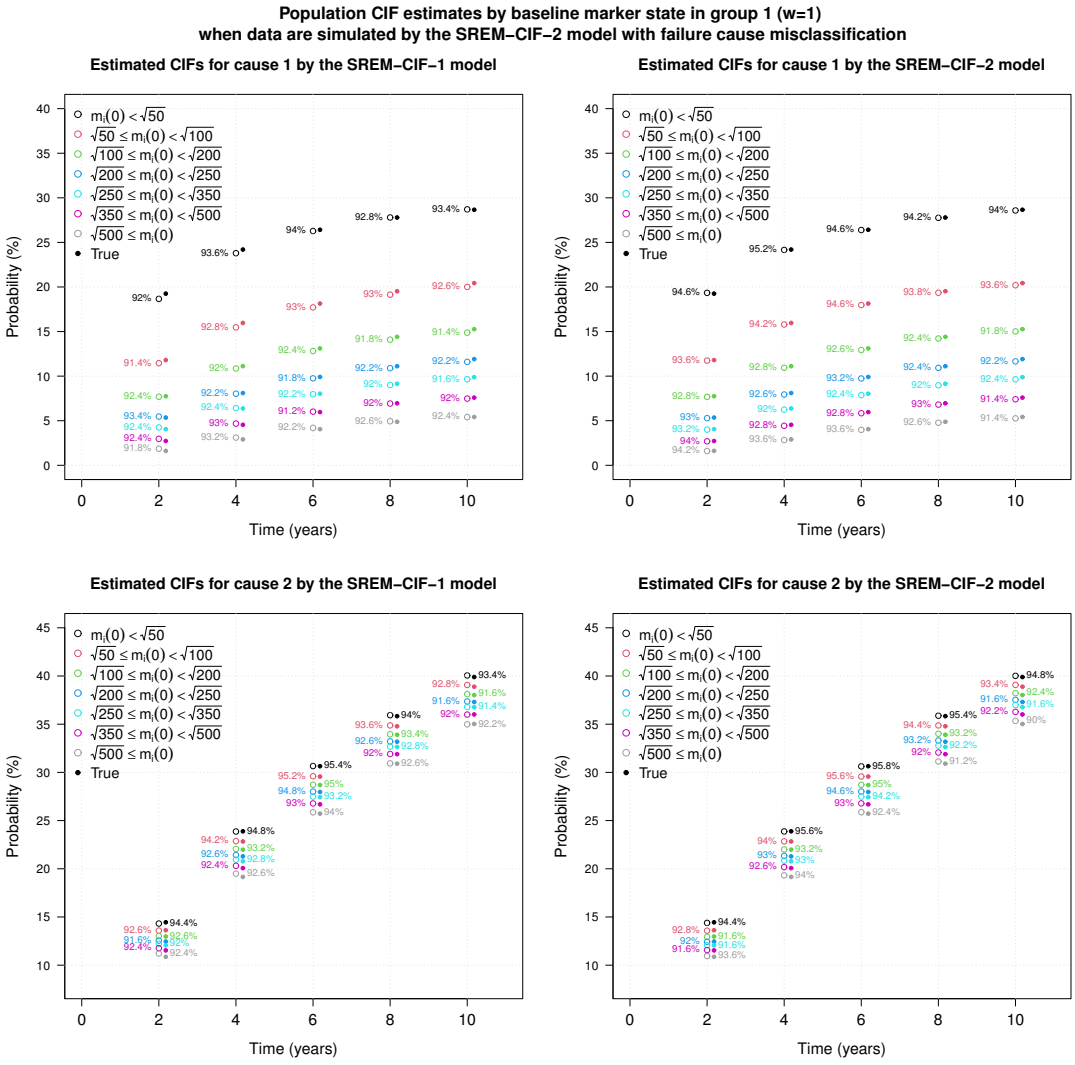


Figure 5.7: Simulation study results for population-averaged CIF estimates by baseline marker state for group 1 when data are simulated under the SREM-CIF-2 model and there is failure cause misclassification. CIF is estimated at certain years since baseline. Open circles show the empirical estimates based on posterior medians over 500 replications whereas closed circles show the true values. Shown are also the corresponding empirical coverage probabilities. The true marker is based on linear splines with knots at 1 and 5 years since baseline and it is correctly specified when fitting SREM-CIF-1 and SREM-CIF-2 models.

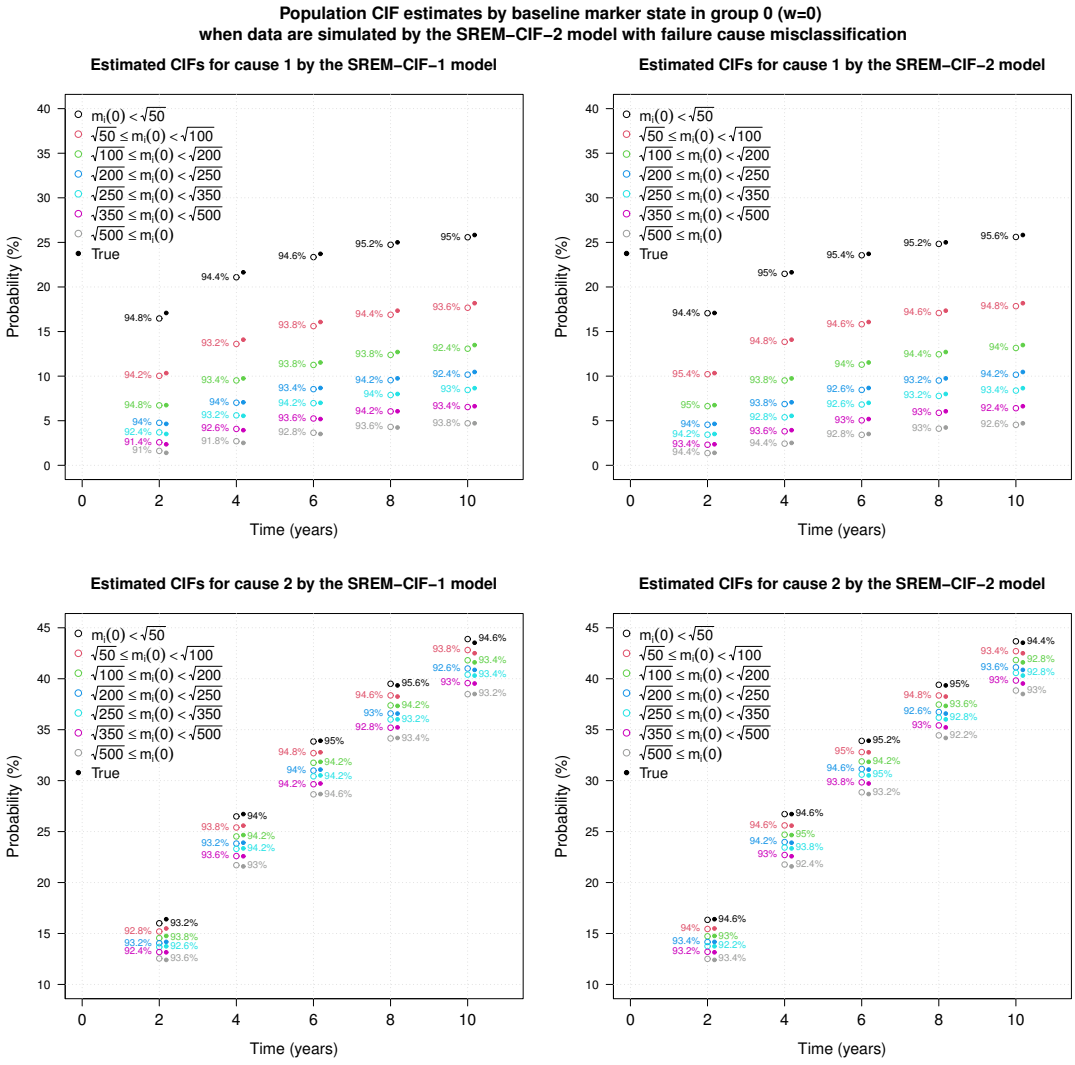


Figure 5.8: Simulation study results for population-averaged CIF estimates by baseline marker state for group 0 when data are simulated under the SREM-CIF-2 model and there is failure cause misclassification. CIF is estimated at certain years since baseline. Open circles show the empirical estimates based on posterior medians over 500 replications whereas closed circles show the true values. Shown are also the corresponding empirical coverage probabilities. The true marker is based on linear splines with knots at 1 and 5 years since baseline and it is correctly specified when fitting SREM-CIF-1 and SREM-CIF-2 models.

5.7 Application

The proposed methodology was applied to data from the East Africa International Epidemiologic Databases to Evaluate AIDS (IeDEA) Regional Consortium. We aimed to jointly estimate the CD4 evolution after ART initiation and the CIFs for death and disengagement from care, assuming that the CIFs depend on the "true" marker values over time. An important issue in the study data is that there may be serious mortality underestimation as many deaths are likely to have been incorrectly classified as disengagements from care. To adjust for that, we incorporated information from double sampling, i.e. a random sample from disengaged patients whose true vital status was ascertained by tracing these patients in the community. It should be noted that, in this application, misclassification can be safely assumed to be uni-directional, that is, a true death can be incorrectly classified as a disengagement from care but a true disengagement from care cannot be misclassified as an observed death. Thus, assuming non-differential misclassification, there is actually only one misclassification parameter π_{11}, with $K = 1$ and $K = 2$ corresponding to death and disengagement from care, respectively. The time gap between two consecutive clinical visits was, on average, approximately equal to one month, with the disengagement date estimated by the the midpoint between the last visit and 2 months after the next scheduled visit date. After taking double sampling data into account, based on the assumed joint model, we estimated multistate probabilities by jointly considering states of "true" CD4 values and the terminal events of death and disengagement from care.

To illustrate the proposed methodology, out of the 61,973 patients with at least one observed CD4 value after ART initiation, we randomly selected 60% of them within each participating clinic, leading to 37,186 patients. To reduce the heterogeneity in the sample, we used data on women aged from 35 to 45 years (the most frequent covariate pattern in the data), leading to 8,005 patients included. The time origin (baseline) was defined at ART initiation, and all available CD4 counts from baseline to the first loss to clinic event were considered. Since there were very few CD4 measurements after

7 years since ART initiation, we censored CD4 observations at 7 years, i.e. all CD4 measurements after 7 years were ignored and the corresponding patients were considered right-censored at 7 years. 3,275 (40.9%) and 273 (3.4%) observed disengagements from care and deaths were seen, respectively, whereas 4,457 (55.7%) were free of any event (right-censored). Deceased patients had substantially lower median (IQR) CD4 counts at baseline [78.0 (31.0, 162.0) cells/μL], compared to disengaged patients [145.0 (67.0, 230.6) cells/μL], and those who were right-censored, 183.0 (99.0, 290.0) cells/μL. In 443 (13.5%) patients out of those who disengaged from care, the true vital status was ascertained through double sampling, in whom there were 80 (18.1%) hidden deaths.

To model the CD4 evolution over time, we used an LMM model on the square root scale using cubic B-splines of time with knots at 0.5, 2, and 4 years since baseline for the fixed effects and a random intercept and slope structure for the random effects. Baseline CIF levels for death and disengagement from care were modeled through cubic B-splines with 2 and 3 knots, respectively. To identify the model with the best fit according to the DIC criterion, we performed a grid search over all combinations of the values $c_1, c_2 \in \{1e-05, 0.25, \ldots, 2\}$ assuming non-differential misclassification π_{11}. We used the prior distributions listed in Subsection 5.2.3, with $\boldsymbol{\mu}_0 = \boldsymbol{\mu}_0^s = \mathbf{0}$, $\boldsymbol{C}_0$ and $\boldsymbol{C}_0^s$ diagonal matrices with variances 100, $\lambda_1 = \lambda_2 = 0.01$, $df = 2$, $\boldsymbol{A} = 2 \times \text{diag}(25, 1)$, and $\pi_{11} \sim \text{Beta}(1, 1)$ (uniform distribution). The DIC criterion was optimized at $c_1 = 1.5$ and $c_2 = 1e - 05$, that is, the subdistribution hazard model yielded the best fit for disengagement from care but not for death. A corresponding joint model assuming that $\pi_{11}(\mathcal{D}_{misc}) = \text{expit}(\boldsymbol{\theta}_{misc,1} + \boldsymbol{\theta}_{misc,2} T_i^\star)$ was also fitted; it was indicated that the probability of correctly classifying a death may decrease at longer times from ART initiation but the effect was not significant, i.e. the odds ratio for one year increase in event time was 0.84 (95%CI: 0.70-1.03). Thus, the final model assumed non-differential misclassification, and for comparison, we also fitted the corresponding model without misclassification. To estimate model parameters, we used 1000 iterations as burn-in, recorded 50000 draws thereafter keeping every tenth posterior value for inferences.

The main results from the two models are presented in Table 5.21. There is sub-

stantial underestimation of mortality, as only 29.21% of the estimated "true" deaths were reported. The fixed effect estimates and the effects of the "true" marker value on the CIF for death were roughly similar between the two models; for example, based on the model accounting for misclassification, the association parameter α_1 for one unit increase in $m_i(t)$ was -0.20 (95%CI: -0.23,-0.17). Of note, the corresponding effects on the CIF of disengagement from care were discordant, i.e. increase of $m_i(t)$ was associated with greater subdistribution hazard for disengagement from care when double sampling data were taken into account, whereas, the model using the observed failure causes implied no effect. This finding may be attributed to the considerable proportion of hidden deaths among the patients flagged as lost to clinic, as deceased patients had lower "true" marker values on average. From now on, we focus only on the model that accounts for misclassification. The estimated CD4 evolution $[(\boldsymbol{x}_i^\top(t)\boldsymbol{\beta})^2]$ is presented in panel A_1 of Figure 5.9. After an initial very rapid increase from ART initiation to around 6 months, CD4 cell counts continued increasing but at a lower rate, until reaching a plateau at about 6 years since ART initiation. The estimates from the proposed model were similar to those obtained by a corresponding LMM (Figure 5.9, panel $\mathbf{A}_1$).

We also estimated the population-averaged CIFs using $L = 1000$ posterior draws keeping every fifth draw, with using $N_{mc} = 1000$. The results are presented in panel A_2 of Figure 5.9. Also shown is the CIFs for death and disengagement from care ignoring potential misclassification, i.e. $\Pr(T_i^\star \leq t, \tilde{K}_i = 1; \boldsymbol{\theta}; \boldsymbol{\theta}_{misc}) = \Pr(T_i^\star \leq t, K_i = 1; \boldsymbol{\theta})\pi_{11}$ and $\Pr(T_i^\star \leq t, \tilde{K}_i = 2; \boldsymbol{\theta}; \boldsymbol{\theta}_{misc}) = \Pr(T_i^\star \leq t, K_i = 1; \boldsymbol{\theta})(1 - \pi_{11}) + \Pr(T_i^\star \leq t, K_i = 2; \boldsymbol{\theta})$, respectively. These estimates are in close agreement with the corresponding estimates using the Aalen–Johansen Estimator, implying that the model is flexible enough to model the observed patterns of events. As expected, CIF estimates ignoring misclassification led to underestimation of mortality and overestimation of the risk for disengagement from care. In panel A_3 of Figure 5.9, we present multistate probabilities over time for all states simultaneously, i.e. both marker states and the terminal events of death and disengagement from care. By the end of the study (7 years since baseline), we estimated that 15.09% (95% CI: 13.13,17.23%) had died

Table 5.21: Results from SREM-CIF-2 models with $c_1 = 1.5$ and $c_2 = 1e-05$ applied to East Africa IeDEA data†.

	Misclassification				No Misclassification			
Parameter	Median	SD	LB	UB	Median	SD	LB	UB
Longitudinal								
Intercept	12.48	0.06	12.35	12.60	12.47	0.06	12.35	12.59
β_1	4.32	0.10	4.13	4.51	4.36	0.10	4.17	4.55
β_2	4.81	0.11	4.59	5.03	4.84	0.11	4.63	5.06
β_3	8.07	0.15	7.78	8.36	8.07	0.15	7.78	8.36
β_4	9.62	0.27	9.10	10.17	9.52	0.28	8.96	10.05
β_5	10.76	0.44	9.88	11.61	10.51	0.44	9.64	11.38
β_6	10.52	0.65	9.25	11.77	10.24	0.66	8.94	11.51
Cause1 (Death)								
"True" marker value, α_1	-0.20	0.01	-0.23	-0.17	-0.18	0.02	-0.22	-0.15
Cause2 (Disengagement)								
"True" marker value sHR, $\exp(\alpha_2)$	1.04	0.01	1.03	1.06	1.00	0.00	0.99	1.01
π_{11}	29.21	1.99	25.56	33.32				

† The mean evolution was based on cubic B-splines of time with knots at 0.5, 2, and 4 years since ART initiation while the random-effects specification was based on a random intercept and slope structure. "Median", "SD", "LB", and "UB" denote the posterior median, standard deviation, 2.5% and 97.5% quantiles, respectively. "sHR" denotes the subdistribution hazard ratio.

and 57.93% (95%CI: 55.22,60.78%) had disengaged from care, whereas, at the end of the study, 0.51% (95%CI: 0.24,0.82%), 0.75% (95%CI: 0.51,1.05%), 2.28% (95%CI: 1.75,2.85%), 1.43% (95%CI: 1.18,1.69%), 3.18% (95%CI: 2.75,3.62%), 4.88% (95%CI: 4.40,5.35%), and 13.95% (95%CI: 11.72,16.09%) were event free with "true" marker states: $m_i(t) < \sqrt{50}$, $\sqrt{50} \leq m_i(t) < \sqrt{100}$, $\sqrt{100} \leq m_i(t) < \sqrt{200}$, $\sqrt{200} \leq m_i(t) < \sqrt{250}$, $\sqrt{250} \leq m_i(t) < \sqrt{350}$, $\sqrt{350} \leq m_i(t) < \sqrt{500}$, $\sqrt{500} \leq m_i(t)$, respectively. The corresponding results for transition probabilities by baseline marker state are presented in Figure 5.10. The CIF of death at 7 years for those starting at $m_i(0) < \sqrt{50}$ was

33.57% (95%CI: 29.85,37.55%), remarkably higher than that of the remaining baseline marker states; even those starting at $\sqrt{50} \leq m_i(0) < \sqrt{100}$ were significantly less likely to die, with the posterior median of the difference being 12.31% (95%CI: 10.30, 14.55%). Conditionally on $m_i(0) < \sqrt{50}$, the transition probability to $m_i(5) \geq \sqrt{500}$ at 5 years while being event-free was as low as 6.93% (95%CI: 5.86,8.01%), whereas the corresponding probability for those with $\sqrt{50} \leq m_i(0) < \sqrt{100}$ was 12.53% (95%CI: 11.36,13.67%).

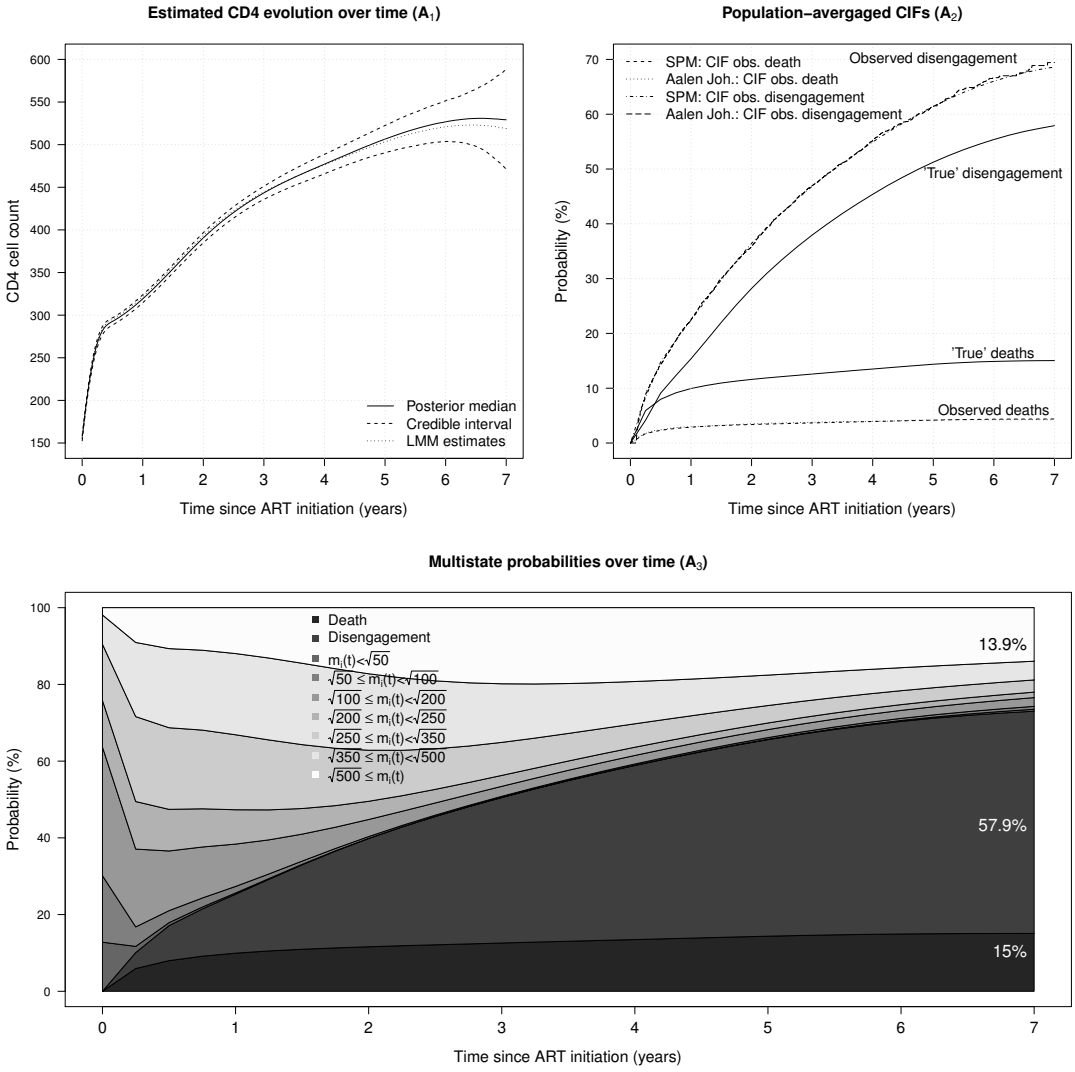


Figure 5.9: Estimated CD4 evolution, population-averaged CIFs, and marker states based on the SREM-CIF-2 model with $c_1 = 1.50$ and $c_2 = 1e-05$, taking the double sampling data into account, applied to East Africa IeDEA data. $\boldsymbol{A}_1$: Estimated CD4 evolution over time since ART initiation, back-transformed on the original scale (from the square-root scale). $\boldsymbol{A}_2$: population-averaged CIFs for death and disengagement from care along with the corresponding CIFs for observed death and disengagement from care. $\boldsymbol{A}_3$: Stacked multistate probability plot of marker states and competing risks for death and disengagement from care over time since ART initiation. The corresponding state occupancy probabilities are visualized through the difference between two adjacent curves with different shades of gray.

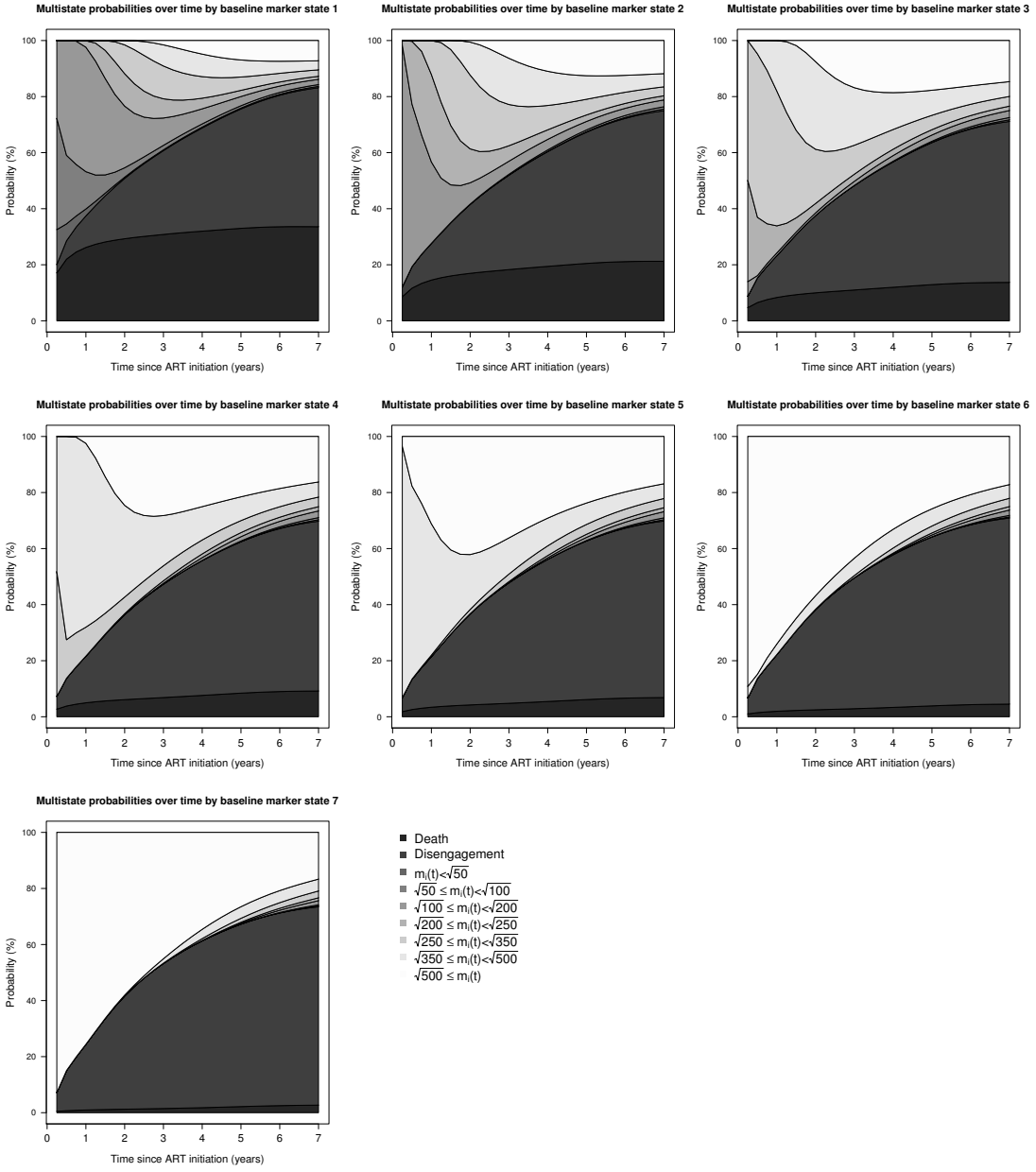


Figure 5.10: Transition probabilities by baseline marker state as estimated by the SREM-CIF-2 model with $c_1 = 1.5$ and $c_2 = 1e-05$ accounting for failure cause misclassification applied to IeDEA data.

5.8 Discussion

In this chapter, we proposed a flexible and unified class of models to jointly model a normally distributed marker over time and competing risks using CIFs for the survival submodels, with inference on model parameters obtained through a hybrid MCMC algorithm. The proposed models assume that the CIFs depend on the "true" marker value over time, $m_i(t)$, thus the association between the marker and survival processes is induced via the random effects. Hence, the proposed models lie within the family of shared random effects models. Most competing-risk shared random effects models rely on cause-specific hazards; but CIF estimates may be of particular interest when the focus lies on population predictions. Though it is feasible to derive CIF estimates based on estimated cause-specific hazards, it requires complex integration, being particularly challenging in joint models. In contrast, under our proposed approach, the effects on the CIFs are described in a direct and straightforward way. To model the link functions, we used the generalized odds rate transformation, with the proportional subdistribution hazards model (Deslandes and Chevret, 2010, Fine and Gray, 1999) being a special case. Due to potential failure cause misclassification in our motivating example, we extended our methodology by incorporating information from doubly sampled patients, i.e. a random sample from patients to whom a gold standard diagnostic procedure was performed. On top of this issue, based solely on the joint model, we also estimated multistate probabilities jointly defined by "true" marker states and competing risks. A simulation study was carried out to examine the performance of the methodology, which indicated that the model performance is satisfactory under certain conditions. The proposed models were also fitted to data from the IeDEA study using CD4 count data from ART initiation until the occurrence of death or disengagement from care. Ignoring double sampling data led to serious underestimation of mortality and implied no effect of the "true" CD4 count, $m_i(t)$, on the risk for disengagement from care, while after adjusting for misclassification positive correlation between $m_i(t)$ and the risk for disengagement from care was revealed. We suppose that this discrepancy could be at

least partly explained by the considerable proportion of deaths among those observed to disengage from care.

One important issue when specifying models for CIFs is that the all-cause CIF should be bounded by 1 at each failure time. We addressed this issue by assuming $\tau_i(\boldsymbol{\beta}, \boldsymbol{\theta}_t, \boldsymbol{b}_i)$ as the upper bound of the survival time, which led to zero likelihood when the constraint is violated. Recall that calculation of $\tau_i(\boldsymbol{\beta}, \boldsymbol{\theta}_t, \boldsymbol{b}_i)$ was not required within the MCMC algorithm, but as noted in Section 5.4, calculation of $\tau_i(\boldsymbol{\beta}, \boldsymbol{\theta}_t, \boldsymbol{b}_i)$ was generally required for valid population-averaged inferences. In summary, we followed Gelfand et al. (1992) who suggested building the constraint in the likelihood function rather than in the prior distribution when the constraint set depends on the data, thus the prior distribution remained $\boldsymbol{b}_i \sim N(\boldsymbol{0}, \boldsymbol{D})$ under our approach.

Our approach of multistate modelling differs from standard approaches (e.g. Putter et al., 2007) in which states are assumed to be directly observed and usually rely on the Markov assumption. In contrast, under our approach, marker states were not assumed to be directly observed, with the computations being solely based on the assumed joint model by formally deriving posterior samples for the quantities of interest, acknowledging that the population CIFs and the marker state probabilities are functions of the parameters of the assumed joint model. Although describing the progression of a cohort via mutually exclusive states over time, defined by both marker and survival data, may offer an alternative view to the data which is frequently used in medical research, relatively few methodological efforts have been made for reliable inference.

It should be noted that the proposed methodology relies on fully parametric models for inference. Hence, as always with parametric approaches, certain assumptions may not hold in practice. Our proposed approach could be extended by assuming more robust measurement errors and potentially heterogeneous random effects (Huang et al., 2010). To obtain greater flexibility, one could also relax the normality assumption on the prior distribution of the random effects by allowing them to have a non-parametric prior distribution through a Dirichlet process (Kleinman and Ibrahim, 1998). Among all model assumptions, though, the ones that are most difficult to verify are perhaps those

related to the missing data mechanisms, e.g. the proposed models, lying within the general class of shared random effects models, assume that missing marker data after the first occurring event are MNAR. However, it has been shown that if the true missing data mechanism causing termination of marker data is MAR and specific SREMs are fitted to such data, the fixed effect estimates are susceptible to bias (Thomadakis et al., 2019). In our application, though, we feel that this is unlikely as the fixed-effect estimates from the proposed model were in line with the corresponding estimates from an analogous LMM.

There are certain extensions that could be incorporated into the proposed methodology. For example, one may also be interested in transition failure probabilities by latent marker state such as $\Pr\{T_i \leq t_2, K_i = k | T_i \geq t_1, m_i(t_1) \in S_{j'}, \boldsymbol{w}_{ik}; \boldsymbol{\theta}\}$, $0 \leq t_1 < t_2$, or in dynamic predictions for specific individuals (Andrinopoulou et al., 2017, Rizopoulos, 2012b). It also needs to be pointed out that, in our application example, the date of disengagement is not exactly known; we instead defined the disengagement date as the one on which patients were expected to no longer have access to medication supplies. Being more precise, the time to disengagement is interval-censored (Bakoyannis et al., 2017), which can be taken into account by adjusting the likelihood for disengaged patients. Given the short time between consecutive visits, though, we expect that the effects of ignoring the interval censoring issue will be minor in this application. Moreover, in the proposed models, the CIFs, or more precisely, some function thereof, are assumed to linearly depend on $m_i(t)$. This assumption could be relaxed by using a more flexible structure.

To sum up, we have proposed a flexible class of shared random effects models to jointly model a normally distributed marker and competing risks using CIFs in the survival submodels. As most approaches in the literature rely on cause-specific hazards, our proposed approach can be a useful alternative when the focus is on identifying risk factors for the risk of an event and on predictions in general. The proposed models have been extended to account for potential failure cause misclassification, and based solely on the assumed joint model, a multistate representation of the whole population

in terms of "true" marker states and competing risks is provided by formally deriving posterior samples for population-averaged CIFs, transitions failure probabilities by baseline marker states, marker state occupation probabilities and transition marker probabilities by baseline states.

Chapter 6

Discussion

Studies that involve longitudinal measurements of a marker related to some chronic disease have become very popular in medical research. The evolution of such markers is typically strongly associated with the evolution of the disease, and thus, longitudinal monitoring of such markers is essential to keep track of disease progression (Jewell and Kalbfleisch, 1992). Moreover, under certain requirements, repeated marker measurements have been used as surrogate outcome measures for hard clinical endpoints in randomized clinical trials set up to evaluate new treatments, especially in diseases with a long time course. For example, in the HIV epidemiology, several immunological and virologic markers have been examined. Among them, the markers that are used more frequently in medical/epidemiological studies are the number of CD4 cells and the HIV-RNA viral load (VL).

Parallel to the extensive use of such longitudinal studies in medical research, plenty of methodological approaches have investigated the consequences on inferences when incomplete marker data are present. The take-home message is that the quality of inferences crucially depends on the nature of the missing data mechanisms, i.e. the probability of a measurement being missing. Nowadays, there are many available methods for modeling longitudinal data with missingness, some of which focus on monotone missingness in the form of dropout. It has been shown that likelihood-based approaches ignoring the dropout mechanism yield consistent estimates given that the

dropout mechanism is MAR and that the likelihood model for the full marker data has been correctly specified. The most difficult case to handle in practice, though, occurs when the missingness probabilities depend on the missing values or on other unobserved characteristics related to the subject-specific marker trajectories (e.g. the random effects). Such missingness mechanisms are referred to as missing not at random (MNAR), whereas the term informative missingness is sometimes adopted, especially for monotone missingness in the form of dropout, which is the focus of this thesis.

There are two main approaches to modeling longitudinal data under MAR missingness, i.e. the use of marginal and the use of conditional models, both reviewed in Chapter 2. The marginal models can be semi-parametrically estimated using the generalized estimating equations (GEE) method, which is a very powerful and versatile semi-parametric approach for marginal inference in longitudinal data. Unfortunately, the appealing properties of GEE hold only under fully observed data or MCAR dropout, i.e. as they do not exploit the full likelihood, they cannot guarantee valid estimates under MAR dropout. One alternative modeling approach is that of the conditional models, more specifically, the generalized linear mixed models (GLMMs). GLMMs are fully parametric approaches, under which, the repeated marker measurements are assumed to be independent conditionally on subject-specific random effects, following the exponential family of distributions. In contrast to marginal models, the regression parameters of GLMMs do not have a marginal interpretation, but rather correspond to the covariate effects on a typical individual. An exception occurs in linear mixed models (LMMs), which correspond to GLMMs with the identity link function along with the Normal distribution for the repeated measurements.

Under MNAR dropout, it is necessary to jointly model the marker measurements and the dropout probabilities to ensure unbiased estimates. There are three main approaches in joint modeling of longitudinal and time-to-dropout data, i.e. (i) selection models, (ii) shared random effects models (SREM), and (iii) pattern-mixture models. These three model frameworks essentially use different factorizations of the full likelihood, i.e. the joint distribution of the full marker data and the dropout mechanism.

In selection models, one specifies a model for the distribution of the full marker data and a (selection) model for the dropout mechanism. The dropout probabilities are assumed to depend on both observed and unobserved marker measurements. The main advantage of selection models is that they can directly estimate the parameters relating to the distribution of the full data, with such estimates usually being of main interest. On the other hand, pattern-mixture models stratify the whole population by pattern of dropout, factorizing the full likelihood as the marker distribution given the dropout pattern and the marginal probabilities of dropout. In SREMs, conceptually, each individual is assumed to have subject-specific random effects influencing both the marker values and the probability of dropout. In theory, SREMs can be considered as either selection models or pattern-mixture models.

One of the first SREM models in the literature was the model proposed by Gruttola and Tu (1994), which was termed LN-SREM(RE) in this thesis. Gruttola and Tu (1994) combined an LMM for the marker model with a log-normal model for the time to dropout, with the mean log-time to dropout assumed to be a linear combination of the random effects plus, possibly, some time-independent covariates. This model, as well as many others (Pantazis et al., 2005, Schluchter, 1992, Touloumi et al., 1999), was developed from a missing data perspective, i.e. to adjust for bias in the estimated marker evolution due to MNAR dropout. Apart from accurately estimating the marker evolution under MNAR dropout, estimation of the risk of dropout conditionally on the "true" (i.e., excluding measurement error) marker value is also of great interest, and it has been shown that simply including the observed marker values in a survival model can severely underestimate the true association (Prentice, 1982, Wulfsohn and Tsiatis, 1997). A standard example of such a modeling approach is the proportional hazards model that includes the "true" current marker value, as predicted by an LMM, in the linear predictor, termed PH-SREM(CV) in this thesis. Most of the literature in joint modeling assumes that there is a single dropout event, but joint modeling of longitudinal data and competing risks has also gained attention.

The research included in this thesis has ultimately been motivated by the epidemi-

ology of HIV infection, mainly concerning the evolution of CD4 cell counts before and after treatment initiation. For example, when studying the CD4 evolution during the HIV natural history (i.e., while individuals remain untreated and AIDS-free), treatment initiation is the primary source of dropout, with the nature of this mechanism (MAR or MCAR) being still debatable (Gras et al., 2013, Pantazis et al., 2005). Assuming MAR dropout, many authors have applied random intercept and slope models to CD4 natural history data. However, if the true covariance structure of the data is more complex, the random intercept and slope model can yield biased results. Therefore, to improve model fit, it has been suggested to either use splines in the design matrix of the random effects (Rizopoulos, 2012b) or to use a random intercept and slope model along with an additional stochastic process such as Brownian motion (BM) (Stirrup et al., 2015, Taylor and Law, 1998).

Focusing on the CD4 evolution after ART initiation, patients on ART may die while in care or disengage from care, which are competing events. Many researchers may be interested in the effects of the "true" (excluding random error) CD4 count on the risk for death and disengagement from care, with this setting calling for joint analysis of longitudinal and competing-risk data. Moreover, in HIV cohort studies, especially in those from resource-constrained countries, substantial under-reporting of deaths is a frequent major issue. That is, patients who have actually died may have been incorrectly classified as disengaged from care, which results in failure cause misclassification. Thus, in this setting, disengagement form care cannot be simply treated simply as non-informative right censoring, as the latter would be valid if individuals who were lost to follow-up were representative of those who were not. In HIV epidemiology, progression of cohorts over time is sometimes monitored by using mutually exclusive states, defined jointly by survival data (e.g. death or disengagement from care) and discretized CD4 data (Stover et al., 2019). In this setting, joint modeling approaches could be advantageous.

Our literate review revealed that the bias of LMMs under MNAR dropout has been well described and recognized and that plenty of SREMs have been proposed including

various parameterizations. Although it is well known that an ultimate discrimination between MAR and MNAR missingness is not possible (Molenberghs et al., 2008), in practice, this seems to have been partially ignored in the joint modeling literature. For example, the effects of fitting an SREM model to MAR data, in terms of induced bias in the parameter estimates, have not been thoroughly studied. Therefore, in Chapter 3, we examined the performance of specific SREM models in estimating the rate of change of longitudinal markers subject to MAR dropout. We analytically calculated the bias in the population slope estimate by both the LN-SREM(RE) and PH-SREM(CV) models. The results revealed that the asymptotic bias in the estimated slope increases as the MAR dropout mechanism strengthens, i.e. at higher MAR dropout probabilities.

In Chapter 3, motivated by the definition of MNAR, we also proposed an alternative model that allows the hazard of dropout to depend on the last observed marker value, as well as on the random effects. In addition, we extended the proposed model to the bivariate case, modeling two correlated markers simultaneously. Based on a simulation study using specific MAR and MNAR dropout scenarios, the proposed model had negligible bias in the estimated slope under both scenarios, whereas the other examined SREM models yielded biased results, especially under the MAR dropout scenario. The examined models were fitted to CD4 data from the CASCADE study, with ART being the dominating reason for dropout. All SREMs, including the proposed one, yielded a steeper CD4 decline compared to the estimate from the LMM, which could indicate presence of an MNAR mechanism, with the SREMs using a random-effects parameterization yielding the most extreme results. In the bivariate version of the proposed model though, that is the one including VL levels as well, the proposed SREM model and the LMM model yielded similar CD4 decline estimates, indicating that, after accounting for viral load levels, the dropout probabilities are unlikely to depend on the underlying CD4 slope. Thus, our bivariate analysis favours the MAR rather than the MNAR assumption for the missing CD4 counts.

In summary, when SREM models are applied in cases where the MAR assumption seems reasonable, results should be interpreted with caution. The proposed model of

Chapter 3 could be alternatively applied when the nature of the dropout mechanism is debatable, at least as part of a sensitivity analysis. Part of this work received a student award in the International Society for Clinical Biostatistics (ISCB) conference, Birmingham (2016), whereas the full paper was published in *Biometrics* (Thomadakis et al., 2019).

Focusing on modeling of marker evolution under MAR dropout, although the claim that an LMM with a stochastic process (e.g. Brownian motion) on top of a simple random-effect structure should be equivalent to an LMM with an elaborate random-effect specification (e.g. through splines) is reasonable, no formal comparison on the basis of bias in parameter estimates under MAR dropout had been evaluated. In Chapter 4, we analytically showed that assuming a simple "random intercept and slope" structure when the true one is substantially more complex can lead to serious biases. We also compared the approach of using a fractional BM process on top of a random intercept and slope structure with the approach of using natural splines for the random effects in terms of their robustness to misspecified covariance structure. For the estimation of a single slope, the two approaches performed almost equally well, though when the "true" population-averaged marker evolution was based on piecewise linear or natural splines, the fractional BM model yielded less biased estimates.

The comparison of the fit of the models in such cases is usually based on the AIC or BIC criteria, as likelihood ratio tests are not possible for non-nested models. However, as the AIC and BIC criteria rely on asymptotics without a straightforward probabilistic interpretation, more formal model comparison criteria could be developed. In Chapter 4, we adopted a Bayesian model comparison approach, based on the posterior model probabilities, to discriminate between LMMs with/without additional BM processes that can have any, and potentially different, random-effect specifications. The examined covariance structure approaches were applied to CD4 natural history data from the CASCADE study. The preferred, by all examined criteria, model was the one including a fractional BM process on top of a random intercept and slope structure. The slope estimate in the simple random intercept and slope model was quite steeper compared

with that from the fractional BM model, which was in line with our analytical findings suggesting biased slope estimates when the fitted covariance structure is simpler than the true one.

To sum up, over-simplistic covariance structures should be avoided under MAR dropout as they could lead to seriously biased fixed-effect estimates. The approach of adding a BM process on top of a random intercept and slope structure or the approach of using an LMM with natural splines for the random effects could be adopted to reduce bias due to potentially misspecified covariance structure. To discriminate between different approaches modeling the covariance structure, the proposed Bayesian model comparison could be applied, especially when there is prior information on the parameters to be estimated. Different parts of this work received a student award in the ninth conference of the Eastern Mediterranean Region of the International Biometrics Society (EMR-IBS), Thessaloniki (2017), and a Best Poster Award in the session Biomedical studies in the ISCB conference, Vigo (2017). The full approach was published in *Statistics in Medicine* (Thomadakis et al., 2020).

Regarding joint modeling of marker data and competing-risk survival data, our literature review revealed that the literature is dominated by cause-specific hazard approaches. However, methods based on cumulative incidence functions (CIFs) are equally valid and may have a more appealing interpretation. In Chapter 5, we proposed a flexible and unified class of models to jointly model a normally distributed marker over time and competing risks using CIFs for the survival submodels. The proposed model assumes that the CIFs depend on the "true" marker value over time, thus the association between the marker and survival processes is induced via the random effects. Hence, the proposed model lies within the family of shared random effects models. Although it is feasible to derive CIF estimates based on estimated cause-specific hazards, it requires complex integration, which is particularly challenging in joint models. In contrast, under the proposed methodology, estimating the CIFs is computationally easier and more direct. Due to death under-reporting in our motivating example, we extended our methodology to take a double sampling design into account (Bakoyannis

et al., 2019), i.e. a small random sample from patients reported to have disengaged from care, in whom, the true vital status is actively ascertained. On top of this issue, based solely on the joint model, we also estimated multistate probabilities jointly defined by "true" marker states and competing risks. A simulation study was carried out to examine the performance of the proposed methodology, which indicated that the model performance is satisfactory under certain conditions. The proposed models were also fitted to data from the IeDEA study using CD4 count data from ART initiation until the occurrence of death or disengagement from care. Ignoring double sampling data led to serious mortality underestimation and implied no effect of the "true" CD4 count on the risk for disengagement from care, while after adjusting for death misclassification, statistically significant positive correlation between the "true" CD4 count and the risk for disengagement from care was found. We suppose that this discrepancy could be at least partly explained by the considerable proportion of deaths among those reported to disengage form care.

One important issue when specifying models for CIFs is that the all-cause CIF should be bounded by 1 at each failure time. We addressed this issue by following Gelfand et al. (1992) who suggested building the constraint in the likelihood function rather than in the prior distribution when the constraint set depends on the data. Our approach of multistate modelling differs from standard approaches (e.g. Putter et al., 2007) in which states are assumed to be directly observed and usually rely on the Markov assumption. In contrast, under our approach, marker states were not assumed to be directly observed, with the computations being solely based on the assumed joint model by formally deriving posterior samples for the quantities of interest, acknowledging that the population CIFs and the marker state probabilities are functions of the parameters of the assumed joint model. The proposed approach could be improved in several ways: e.g. by including more robust measurement errors or potentially heterogeneous random effects (Huang et al., 2010). As with all parametric models, the validity of the estimates depend on the correctness of model assumptions. Among all model assumptions, though, the ones that are most difficult to verify in practice are perhaps

those related to the missing data mechanisms. Recall that the proposed approach lies within the SREM class of MNAR models. However, in Chapter 3, we showed that if the true dropout mechanism causing termination of marker data is MAR and specific SREMs are fitted to such data, the fixed-effect estimates are susceptible to bias. In this application, though, we feel that this is unlikely as the fixed-effect estimates from the proposed model were in line with the corresponding estimates from the LMM.

To summarize, as most approaches in the literature rely on cause-specific hazards, our proposed approach can be a useful alternative when one focuses on identifying risks factors for the risk of an event and on predictions in general. The proposed model has been extended to account for potential failure cause misclassification through double sampling, as in HIV cohort studies it is highly likely that the population of disengaged patients includes hidden (unobserved) deaths in care. Moreover, we proposed a unified and formal approach to estimate state occupation and transition probabilities in terms of "true" or observed marker states and competing risks based solely on the assumed SREM joint model. Parts of this work were presented as oral presentations in the ISCB conference, Leuven (2019) and Lyon (2021). The full paper and has submitted to *Biostatistics* and is currently under revision.

To summarize, this thesis focuses mainly on joint modelling of a disease marker evolution and clinical outcomes. Its contribution can be summarized as follows:

1. It has been shown that SREMs under MAR marker data due to dropout could lead to seriously biased marker evolution estimates. An alternative model has been proposed, which could be used, at least as a sensitivity analysis, when the nature of the drop out mechanism is not known, with its performance evaluated through simulation studies. The proposed model performs well under specific MAR and MNAR missing marker data.

2. When fitting LMMs under MAR dropout, it has been shown that a simple covariance structure when the true one is more complex can lead to seriously biased fixed-effect estimates. To minimize induced biases, alternative approaches, such

including a BM process or including splines in the design matrix of random effects, have been evaluated. In addition, a Bayesian model comparison criterion has been proposed to compare non-nested models, e.g. when models with different covariance structure are compared.

3. A unified method to model the evolution of a disease marker along with its effect on the CIFs of the competing events, subject also to failure cause misclassification, has been proposed and evaluated through simulation studies. The proposed method can be used: a) to estimate the effect of the underlying marker level on the risk of each of the competing outcomes; b) to estimate the probability that a randomly chosen individual from the population is in specific states, defined in terms of underlying or observed marker level and clinical outcomes. Transition probabilities between states have also been derived. For these approaches, several methodological/statistical issues have been dealt with. The proposed unified approach could be applied not only in infectious diseases (such as the HIV, which was used as an example in this thesis) but also in chronic diseases (such as cardiovascular diseases and cancer studies) for predictions purposes or for evaluating public health interventions.

Bibliography

T. Amemiya. Tobit models: A survey. *Journal of Econometrics*, 24(1-2):3–61, 1984. URL https://EconPapers.repec.org/RePEc:eee:econom:v:24:y:1984:i:1-2:p:3-61. 44

E.-R. Andrinopoulou, D. Rizopoulos, J. J. M. Takkenberg, and E. Lesaffre. Joint modeling of two longitudinal outcomes and competing risk data. *Statistics in Medicine*, 33(18):3167–3178, 2014. doi: 10.1002/sim.6158. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.6158. 6, 55, 56, 57

E.-R. Andrinopoulou, D. Rizopoulos, J. J. Takkenberg, and E. Lesaffre. Combined dynamic predictions using joint models of two longitudinal outcomes and competing risk data. *Statistical Methods in Medical Research*, 26(4):1787–1801, 2017. doi: 10.1177/0962280215588340. URL https://doi.org/10.1177/0962280215588340. PMID: 26059114. 6, 55, 56, 57, 229

G. Bakoyannis and G. Touloumi. Practical methods for competing risks data: A review. *Statistical Methods in Medical Research*, 21(3):257–272, 2012. doi: 10.1177/0962280210394479. URL https://doi.org/10.1177/0962280210394479. PMID: 21216803. 56, 57

G. Bakoyannis and C. T. Yiannoutsos. Impact of and correction for outcome misclassification in cumulative incidence estimation. *PLOS ONE*, 10(9):1–15, 09 2015. doi: 10.1371/journal.pone.0137454. URL https://doi.org/10.1371/journal.pone.0137454. 171

BIBLIOGRAPHY

G. Bakoyannis, M. Yu, and C. T. Yiannoutsos. Semiparametric regression on cumulative incidence function with interval-censored competing risks data. *Statistics in Medicine*, 36(23):3683–3707, 2017. doi: 10.1002/sim.7350. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.7350`. 57, 58, 160, 229

G. Bakoyannis, Y. Zhang, and C. T. Yiannoutsos. Nonparametric inference for markov processes with missing absorbing state. *Statistica Sinica*, 29(4):2083–2104, 2019. 59, 169, 237

E. Balestre, S. P. Eholié, A. Lokossue, P. S. Sow, M. Charurat, A. Minga, J. Drabo, F. Dabis, D. K. Ekouevi, R. Thiébaut, et al. Effect of age on immunological response in the first year of antiretroviral therapy in hiv-1-infected adults in west africa. *AIDS*, 26(8):951–957, 2012. 9

J. Barnard, R. McCulloch, and X. Meng. Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, 10(4):1281–1311, 10 2000. ISSN 1017-0405. 142, 143, 144, 145, 146

J. Bates, J. Pinheiro, J. Pinheiro, and D. Bates. *Mixed-Effects Models in S and S-PLUS*. Statistics and Computing. Springer New York, 2000. ISBN 9780387989570. URL `https://books.google.gr/books?id=N3WeyHFbHLQC`. 39, 40

J. Beyersmann, A. Allignol, and M. Schumacher. *Competing risks and multistate models with R*. Springer Science & Business Media, 2011. 56, 171, 182

P. J. Bickel and J. Kwon. Inference for semiparametric models: Some questions and an answer. *Statistica Sinica*, 11(4):863–886, 2001. ISSN 10170405, 19968507. URL `http://www.jstor.org/stable/24306883`. 35

N. E. Breslow and D. G. Clayton. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88(421):9–25, 1993. ISSN 01621459. URL `http://www.jstor.org/stable/2290687`. 36, 37

E. R. Brown, J. G. Ibrahim, and V. DeGruttola. A flexible b-spline model for multiple longitudinal biomarkers and survival. *Biometrics*, 61(1):64–73, 2005. doi: 10.1111/j.0006-341X.2005.030929.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.0006-341X.2005.030929.x. 41

M. Capanu, M. Gönen, and C. B. Begg. An assessment of estimation methods for generalized linear mixed models with binary outcomes. *Statistics in Medicine*, 32(26):4550–4566, 2013. doi: 10.1002/sim.5866. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.5866. 37

J. R. Carpenter, M. G. Kenward, and S. Vansteelandt. A comparison of multiple imputation and doubly robust estimation for analyses with missing data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169(3):571–584, 2006. doi: 10.1111/j.1467-985X.2006.00407.x. URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-985X.2006.00407.x. 35

J. B. Cavender, W. J. Rogers, L. D. Fisher, B. J. Gersh, C. J. Coggin, and W. O. Myers. Effect of smoking on survival and morbidity in patients randomized to medical or surgical therapy in the coronary artery surgery study (cass): 10-year follow-up. *Journal of the American College of Cardiology*, 20(2):287–294, 1992. 51

Y.-Y. Chi and J. G. Ibrahim. Joint models for multivariate longitudinal and multivariate survival data. *Biometrics*, 62(2):432–445, 2006. ISSN 0006341X, 15410420. URL http://www.jstor.org/stable/3695862. 53

S. Choi, S. W. Lagakos, R. T. Schooley, and P. A. Volberding. Cd4+ lymphocytes are an incomplete surrogate marker for clinical progression in persons with asymptomatic hiv infection taking zidovudine. *Annals of Internal Medicine*, 118(9):674–680, 1993. doi: 10.7326/0003-4819-118-9-199305010-00003. URL https://www.acpjournals.org/doi/abs/10.7326/0003-4819-118-9-199305010-00003. PMID: 8096373. 8

D. Christophe and S. Petr. *randtoolbox: Generating and Testing Random Numbers*, 2019. R package version 1.30.0. 81

C. Collaboration. Changes in the uptake of antiretroviral therapy and survival in people with known duration of hiv infection in europe: results from cascade. *HIV Medicine*, 1(4):224–231, 2000. doi: 10.1046/j.1468-1293.2000.00033.x. URL `https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1468-1293.2000.00033.x`. 10

T. D. Cook and M. R. Kosorok. Analysis of time-to-event data with incomplete event adjudication. *Journal of the American Statistical Association*, 99(468):1140–1152, 2004. ISSN 01621459. URL `http://www.jstor.org/stable/27590492`. 59

J. Copas and S. Eguchi. Local sensitivity approximations for selectivity bias. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(4):871–895, 2001. doi: 10.1111/1467-9868.00318. URL `https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/1467-9868.00318`. 23

A. Creemers, N. Hens, M. Aerts, G. Molenberghs, G. Verbeke, and M. G. Kenward. Generalized shared-parameter models and missingness at random. *Statistical Modelling*, 11(4):279–310, 2011. doi: 10.1177/1471082X1001100401. URL `https://doi.org/10.1177/1471082X1001100401`. 23, 47

M. J. Crowther. STJM: Stata module to fit shared parameter joint models of longitudinal and survival data. Statistical Software Components, Boston College Department of Economics, Aug. 2012. URL `https://ideas.repec.org/c/boc/bocode/s457502.html`. 54, 99

M. J. Crowther, K. R. Abrams, and P. C. Lambert. Flexible parametric joint modelling of longitudinal and survival data. *Statistics in Medicine*, 31(30):4456–4471, 2012. doi: 10.1002/sim.5644. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.5644`. 53, 55, 83

M. J. Crowther, T. M.-L. Andersson, P. C. Lambert, K. R. Abrams, and K. Humphreys. Joint modelling of longitudinal and survival data: incorporating delayed entry and an assessment of model misspecification. *Statistics in Medicine*, 35(7):1193–1209,

2016. doi: 10.1002/sim.6779. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.6779. 53, 55

D. M. Dabrowska and K. A. Doksum. Estimation and testing in a two-sample generalized odds-rate model. *Journal of the American Statistical Association*, 83(403): 744–749, 1988. ISSN 01621459. URL http://www.jstor.org/stable/2289300. 160

U. G. Dafni and A. A. Tsiatis. Evaluating surrogate markers of clinical outcome when measured with error. *Biometrics*, 54(4):1445–1462, 1998. ISSN 0006341X, 15410420. URL http://www.jstor.org/stable/2533670. 54

C. Daniel Paulino, P. Soares, and J. Neuhaus. Binomial regression with misclassification. *Biometrics*, 59(3):670–675, 2003. doi: 10.1111/1541-0420.00077. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1541-0420.00077. 59, 169, 171

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977. ISSN 00359246. URL http://www.jstor.org/stable/2984875. 40, 45

E. Deslandes and S. Chevret. Joint modeling of multivariate longitudinal data and the dropout process in a competing risk setting: application to icu data. *BMC Medical Research Methodology*, 10:69, 2010. 56, 161, 227

R. Detels, P. English, J. Giorgi, B. Visscher, J. Fahey, J. Taylor, J. Dudley, P. Nishanian, A. Muñoz, and J. Phair. Patterns of cd4+ cell changes after hiv-1 infection indicate the existence of a codeterminant of aids. *Journal of acquired immune deficiency syndromes*, 1(4):390—395, 1988. ISSN 0894-9255. URL http://europepmc.org/abstract/MED/2905742. 8

P. Diggle and M. G. Kenward. Informative drop-out in longitudinal data analysis. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 43(1):49–93, 1994. ISSN 00359254, 14679876. URL http://www.jstor.org/stable/2986113. 5, 44, 45, 46

P. Diggle, D. Diggle, F. allgemeine tierzucht, O. U. Press, P. Diggle, P. Heagerty, K. Liang, P. Heagerty, S. Zeger, and B. Zeger. *Analysis of Longitudinal Data*. Oxford Statistical Science Series. OUP Oxford, 2002. ISBN 9780198524847. URL https://books.google.gr/books?id=kKLbyWycRwcC. 2, 25, 29, 30

R. Drikvandi, G. Verbeke, A. Khodadadi, and V. Partovi Nia. Testing multiple variance components in linear mixed-effects models. *Biostatistics*, 14(1):144–159, 08 2012. ISSN 1465-4644. doi: 10.1093/biostatistics/kxs028. URL https://doi.org/10.1093/biostatistics/kxs028. 115

R. M. Elashoff, G. Li, and N. Li. A joint model for longitudinal measurements and survival data in the presence of multiple failure types. *Biometrics*, 64(3):762–771, 2008. doi: 10.1111/j.1541-0420.2007.00952.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1541-0420.2007.00952.x. 6, 55, 56, 57

M. Eyster, M. Gail, J. Ballard, H. Al-Mondhiry, and J. Goedert. Natural history of human immunodeficiency virus infections in hemophiliacs: effects of t-cell subsets, platelet counts, and age. *Annals of internal medicine*, 107(1):1—6, July 1987. ISSN 0003-4819. doi: 10.7326/0003-4819-107-1-1. URL https://doi.org/10.7326/0003-4819-107-1-1. 8

J. L. Fahey, J. M. Taylor, R. Detels, B. Hofmann, R. Melmed, P. Nishanian, and J. V. Giorgi. The prognostic value of cellular and serologic markers in infection with human immunodeficiency virus type 1. *New England Journal of Medicine*, 322(3):166–172, 1990. doi: 10.1056/NEJM199001183220305. URL https://doi.org/10.1056/NEJM199001183220305. PMID: 1967191. 8

D. Farewell, R. Daniel, and S. Seaman. Missing at random: a stochastic process perspective. *arXiv preprint arXiv:1801.06739*, 2018. 154

C. L. Faucett and D. C. Thomas. Simultaneously modelling censored survival data and repeatedly measured covariates: a gibbs sampling approach. *Statistics in Medicine*, 15(15):1663–1685, 1996. doi:

10.1002/(SICI)1097-0258(19960815)15:15⟨1663::AID-SIM294⟩3.0.CO;2-1. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291097-0258%2819960815%2915%3A15%3C1663%3A%3AAID-SIM294%3E3.0.CO%3B2-1. 5, 6, 50

J. P. Fine and R. J. Gray. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, 94(446):496–509, 1999. ISSN 01621459. URL http://www.jstor.org/stable/2670170. 57, 58, 161, 227

G. Fitzmaurice, M. Davidian, G. Verbeke, and G. Molenberghs. *Longitudinal Data Analysis*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. CRC Press, 2008. ISBN 9781420011579. 3, 25

G. Fitzmaurice, N. Laird, and J. Ware. *Applied Longitudinal Analysis*. Wiley Series in Probability and Statistics. Wiley, 2011. ISBN 9780470380277. URL https://books.google.gr/books?id=qOmxRtdNJpEC. 2, 25

D. Gamerman. Sampling from the posterior distribution in generalized linear mixed models. *Statistics and Computing*, 7:57–68, 1997. 164, 173

A. García-Hernandez, T. Pérez, M. d. C. Pardo, and D. Rizopoulos. Mmrm vs joint modeling of longitudinal responses and time to study drug discontinuation in clinical trials using a "de jure" estimand. *Pharmaceutical Statistics*, n/a(n/a), 2020. doi: 10.1002/pst.2045. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/pst.2045. 28, 53

A. E. Gelfand, A. F. M. Smith, and T.-M. Lee. Bayesian analysis of constrained parameter and truncated data problems using gibbs sampling. *Journal of the American Statistical Association*, 87(418):523–532, 1992. ISSN 01621459. URL http://www.jstor.org/stable/2290286. 58, 228, 238

P. Gilbert and R. Varadhan. *numDeriv: Accurate Numerical Derivatives*, 2016. URL https://CRAN.R-project.org/package=numDeriv. R package version 2016.8-1. 79

L. Gras, R. B. Geskus, S. Jurriaans, M. Bakker, A. van Sighem, D. Bezemer, et al. Has the rate of cd4 cell count decline before initiation of antiretroviral therapy changed over the course of the dutch hiv epidemic among msm? *PLoS One*, 8(5):e64437, 2013. 12, 62, 234

G. Grimmett, G. Grimmett, D. Stirzaker, P. Grimmett, and M. Stirzaker. *Probability and Random Processes*. Probability and Random Processes. OUP Oxford, 2001. ISBN 9780198572220. URL `https://books.google.gr/books?id=G3ig-0M4wSIC`. 41

V. D. Gruttola and X. M. Tu. Modelling progression of cd4-lymphocyte count and its relationship to survival time. *Biometrics*, 50(4):1003–1014, 1994. ISSN 0006341X, 15410420. URL `http://www.jstor.org/stable/2533439`. 5, 6, 49, 65, 233

C. Han and B. P. Carlin. Markov chain monte carlo methods for computing bayes factors: A comparative review. *Journal of the American Statistical Association*, 96(455):1122–1132, 2001. ISSN 01621459. URL `http://www.jstor.org/stable/2670258`. 115

T. E. Hanson, A. J. Branscum, and W. O. Johnson. Predictive comparison of joint longitudinal-survival modeling: a case study illustrating competing approaches. *Lifetime data analysis*, 17(1):3–28, 2011. 53

D. A. Harville. Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, 72(358):320–338, 1977. ISSN 01621459. URL `http://www.jstor.org/stable/2286796`. 2, 3, 36, 39

D. A. Harville. *Matrix Algebra From a Statistician's Perspective*. Springer, New York, 1997. 68, 72, 73, 143

D. A. Harville and R. W. Mee. A mixed-model procedure for analyzing ordered categorical data. *Biometrics*, 40(2):393–408, 1984. ISSN 0006341X, 15410420. URL `http://www.jstor.org/stable/2531393`. 3

J. J. Heckman. The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models. In *Annals of Economic and Social Measurement, Volume 5, number 4*, NBER Chapters, pages 475–492. National Bureau of Economic Research, Inc, July 1976. URL `https://ideas.repec.org/h/nbr/nberch/10491.html`. 44

R. Henderson, P. Diggle, and A. Dobson. Joint modelling of longitudinal measurements and event time data . *Biostatistics*, 1(4):465–480, 12 2000. ISSN 1465-4644. doi: 10.1093/biostatistics/1.4.465. URL `https://doi.org/10.1093/biostatistics/1.4.465`. 53, 55

G. L. Hickey, P. Philipson, A. Jorgensen, and R. Kolamunnage-Dona. A comparison of joint models for longitudinal and competing risks data, with application to an epilepsy drug randomized controlled trial. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 181(4):1105–1123, 2018. doi: 10.1111/rssa.12348. URL `https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/rssa.12348`. 6, 55

M. Holodniy, D. A. Katzenstein, S. Sengupta, A. M. Wang, C. Casipit, D. H. Schwartz, M. Konrad, E. Groves, and T. C. Merigan. Detection and Quantification of Human Immunodeficiency Virus RNA in Patient Serum by Use of the Polymerase Chain Reaction. *The Journal of Infectious Diseases*, 163(4):862–866, 04 1991. ISSN 0022-1899. doi: 10.1093/infdis/163.4.862. URL `https://doi.org/10.1093/infdis/163.4.862`. 9

D. G. Horvitz and D. J. Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47 (260):663–685, 1952. doi: 10.1080/01621459.1952.10483446. URL `https://amstat.tandfonline.com/doi/abs/10.1080/01621459.1952.10483446`. 5, 34

F. Hsieh, Y.-K. Tseng, and J.-L. Wang. Joint modeling of survival and longitudi-

nal data: Likelihood approach revisited. *Biometrics*, 62(4):1037–1043, 2006. ISSN 0006341X, 15410420. URL http://www.jstor.org/stable/4124524. 53, 55

B. Hu, L. Li, X. Wang, and T. Greene. Nonparametric multistate representations of survival and longitudinal data with measurement error. *Statistics in Medicine*, 31(21):2303–2317, 2012. doi: 10.1002/sim.5369. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.5369. 6, 7, 59

W. Hu, G. Li, and N. Li. A bayesian approach to joint analysis of longitudinal measurements and competing risks failure time data. *Statistics in Medicine*, 28(11): 1601–1619, 2009. doi: 10.1002/sim.3562. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.3562. 6, 55, 56

X. Huang, G. Li, and R. M. Elashoff. A joint model of longitudinal and competing risks survival data with heterogeneous random effects and outlying longitudinal measurements. *Statistics and its interface*, 3:185–195, 07 2010. 5, 6, 55, 111, 228, 238

J. G. Ibrahim, M. H. Chen, and D. Sinha. *Bayesian survival analysis*. Springer, 2001. ISBN 9783540952770. URL https://books.google.gr/books?id=flVcSQAACAAJ. 94

I. Jansen and G. Molenberghs. A flexible marginal modelling strategy for non-monotone missing data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171(2):347–373, 2008. doi: 10.1111/j.1467-985X.2007.00524.x. URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-985X.2007.00524.x. 47

J.-H. Jeong and J. P. Fine. Parametric regression on cumulative incidence function. *Biostatistics*, 8(2):184–196, 04 2006. ISSN 1465-4644. doi: 10.1093/biostatistics/kxj040. URL https://doi.org/10.1093/biostatistics/kxj040. 58, 160, 161

N. P. Jewell and J. D. Kalbfleisch. *Marker Models in Survival Analysis and Applications to Issues Associated with AIDS*, pages 211–230. Birkhäuser Boston, Boston, MA,

1992. ISBN 978-1-4757-1229-2. doi: 10.1007/978-1-4757-1229-2_10. URL https://doi.org/10.1007/978-1-4757-1229-2_10. 1, 231

Joint United Nations Programme on HIV/AIDS and Joint United Nations Programme on HIV/Aids. 90-90-90: an ambitious treatment target to help end the aids epidemic. *Geneva: Unaids*, 2014. URL https://www.unaids.org/en/resources/documents/2017/90-90-90. 11

R. H. Jones. Bayesian information criterion for longitudinal and clustered data. *Statistics in Medicine*, 30(25):3050–3056, 2011. doi: 10.1002/sim.4323. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.4323. 147, 148, 152

S. Jurriaans, B. Van Gemen, G. Weverling, D. Van Strijp, P. Nara, R. Coutinho, M. Koot, H. Schuitemaker, and J. Goudsmit. The natural history of hiv-1 infection: virus load and virus phenotype independent determinants of clinical course? *Virology*, 204(1):223—233, October 1994. ISSN 0042-6822. doi: 10.1006/viro.1994.1526. URL https://doi.org/10.1006/viro.1994.1526. 9

J. D. Kalbfleisch and R. L. Prentice. *The Statistical Analysis of Failure Time Data*, volume 360. John Wiley & Sons, 2002. 54, 56

R. E. Kass and A. E. Raftery. Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795, 1995. ISSN 01621459. URL http://www.jstor.org/stable/2291091. 115

M. G. Kenward. Selection models for repeated measurements with non-random dropout: an illustration of sensitivity. *Statistics in Medicine*, 17(23):2723–2732, 1998. doi: 10.1002/(SICI)1097-0258(19981215)17:23⟨2723::AID-SIM38⟩3.0.CO;2-5. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291097-0258%2819981215%2917%3A23%3C2723%3A%3AAID-SIM38%3E3.0.CO%3B2-5. 46

H. J. Keselman, J. Algina, R. K. Kowalchuk, and R. D. Wolfinger. A comparison of two approaches for selecting covariance structures in the analysis of repeated

measurements. *Communications in Statistics - Simulation and Computation*, 27(3): 591–604, 1998. doi: 10.1080/03610919808813497. URL https://doi.org/10.1080/03610919808813497. 115, 153

S. Kim, D. Zeng, and J. M. G. Taylor. Joint partially linear model for longitudinal data with informative drop-outs. *Biometrics*, 73(1):72–82, 2017. doi: 10.1111/biom.12566. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/biom.12566. 55

J. P. Klein and Y. Shu. Multi-state models for bone marrow transplantation studies. *Statistical Methods in Medical Research*, 11(2):117–139, 2002. doi: 10.1191/0962280202sm277ra. URL https://doi.org/10.1191/0962280202sm277ra. PMID: 12040693. 7

K. P. Kleinman and J. G. Ibrahim. A semiparametric bayesian approach to the random effects model. *Biometrics*, 54(3):921–938, 1998. ISSN 0006341X, 15410420. URL http://www.jstor.org/stable/2533846. 228

N. M. Laird. Missing data in longitudinal studies. *Statistics in Medicine*, 7(1-2):305–315, 1988. doi: 10.1002/sim.4780070131. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.4780070131. 4, 5, 18, 21, 22, 23

N. M. Laird and J. H. Ware. Random-effects models for longitudinal data. *Biometrics*, 38(4):963–974, 1982. ISSN 0006341X, 15410420. URL http://www.jstor.org/stable/2529876. 2, 3, 36, 39, 158

P. C. Lambert and P. Royston. Further development of flexible parametric models for survival analysis. *The Stata Journal*, 9(2):265–290, 2009. doi: 10.1177/1536867X0900900206. URL https://doi.org/10.1177/1536867X0900900206. 120

W. Lang, H. Perkins, R. E. Anderson, R. Royce, N. Jewell, and W. Winkelstein Jr. Patterns of t lymphocyte changes with human immunodeficiency virus infection: from seroconversion to the development of aids. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, 2(1):63–69, 1989. 8

T. Leonard and J. S. J. Hsu. Bayesian inference for a covariance matrix. *The Annals of Statistics*, 20(4):1669–1696, 1992. ISSN 00905364. URL http://www.jstor.org/stable/2242363. 40, 142

K.-Y. Liang and S. L. Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22, 1986. ISSN 00063444. URL http://www.jstor.org/stable/2336267. 2, 3, 32, 33

K.-Y. Liang, S. L. Zeger, and B. Qaqish. Multivariate regression analyses for categorical data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 54(1):3–40, 1992. ISSN 00359246. URL http://www.jstor.org/stable/2345947. 33

M. J. Lindstrom and D. M. Bates. Newton-raphson and em algorithms for linear mixed-effects models for repeated-measures data. *Journal of the American Statistical Association*, 83(404):1014–1022, 1988. ISSN 01621459. URL http://www.jstor.org/stable/2290128. 3, 40, 68, 71, 74

S. R. Lipsitz, N. M. Laird, and D. P. Harrington. Generalized estimating equations for correlated binary data: Using the odds ratio as a measure of association. *Biometrika*, 78(1):153–160, 1991. ISSN 00063444. URL http://www.jstor.org/stable/2336905. 33

S. R. Lipsitz, N. M. Laird, and D. P. Harrington. A three-stage estimator for studies with repeated and possibly missing binary outcomes. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 41(1):203–213, 1992. ISSN 00359254, 14679876. URL http://www.jstor.org/stable/2347629. 33

S. R. Lipsitz, G. Molenberghs, G. M. Fitzmaurice, and J. Ibrahim. Gee with gaussian estimation of the correlations when data are incomplete. *Biometrics*, 56(2):528–536, 2000. doi: 10.1111/j.0006-341X.2000.00528.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.0006-341X.2000.00528.x. 33

R. Little and D. Rubin. *Statistical Analysis with Missing Data*. Wiley Series in Prob-

ability and Statistics. Wiley, 2002. ISBN 9780471183860. URL https://books.google.gr/books?id=aYPwAAAAMAAJ. 4, 5, 21, 23

R. J. A. Little. Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association*, 88(421):125–134, 1993. ISSN 01621459. URL http://www.jstor.org/stable/2290705. 5, 60

R. J. A. Little. A class of pattern-mixture models for normal incomplete data. *Biometrika*, 81(3):471–483, 1994. ISSN 00063444. URL http://www.jstor.org/stable/2337120. 5, 60

R. J. A. Little. Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association*, 90(431):1112–1121, 1995. doi: 10.1080/01621459.1995.10476615. URL https://www.tandfonline.com/doi/abs/10.1080/01621459.1995.10476615. 19, 46

R. J. A. Little and Y. Wang. Pattern-mixture models for multivariate incomplete data with covariates. *Biometrics*, 52(1):98–111, 1996. ISSN 0006341X, 15410420. URL http://www.jstor.org/stable/2533148. 60

K. Lu and D. V. Mehrotra. Specification of covariance structure in longitudinal data analysis for randomized clinical trials. *Statistics in Medicine*, 29(4):474–488, 2010. doi: 10.1002/sim.3820. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.3820. 28, 114

J. D. Lundgren, A. G. Babiker, F. Gordin, S. Emery, B. Grund, S. Sharma, A. Avihingsanon, D. A. Cooper, G. Fätkenheuer, J. M. Llibre, et al. Initiation of antiretroviral therapy in early asymptomatic hiv infection. *New England Journal of Medicine*, 373(9):795–807, 2015. doi: 10.1056/NEJMoa1506816. URL https://doi.org/10.1056/NEJMoa1506816. PMID: 26192873. 9

R. Lyles, L. Tang, H. Superak, C. King, D. Celentano, Y. Lo, and J. Sobel. Validation data-based adjustments for outcome misclassification in logistic regression:

An illustration. *Epidemiology*, 22(4):589–598, jul 2011. ISSN 1044-3983. doi: 10.1097/EDE.0b013e3182117c85. 59

R. H. Lyles, A. Muñoz, T. E. Yamashita, H. Bazmi, R. Detels, C. R. Rinaldo, J. B. Margolick, J. P. Phair, and f. t. M. A. C. S. Mellors, John W. Natural History of Human Immunodeficiency Virus Type 1 Viremia after Seroconversion and Proximal to AIDS in a Large Cohort of Homosexual Men. *The Journal of Infectious Diseases*, 181(3):872–880, 03 2000. ISSN 0022-1899. doi: 10.1086/315339. URL `https://doi.org/10.1086/315339`. 9

J. R. Magnus. On differentiating eigenvalues and eigenvectors. *Econometric Theory*, 1 (2):179–191, 1985. 143

B. B. Mandelbrot and J. W. V. Ness. Fractional brownian motions, fractional noises and applications. *SIAM Review*, 10(4):422–437, 1968. ISSN 00361445. URL `http://www.jstor.org/stable/2027184`. 42

C. F. Manski. *Analog Estimation Methods in Econometrics*. Monographs on Statistics & Applied Probability. Chapman and Hall, 1988. ISBN 9780412011412. URL `https://books.google.gr/books?id=OyO7AAAAIAAJ`. 65

L. Mao and D. Y. Lin. Efficient estimation of semiparametric transformation models for the cumulative incidence of competing risks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(2):573–587, 2017. doi: 10.1111/rssb.12177. URL `https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/rssb.12177`. 57, 58

K. Mauff, E. Steyerberg, I. Kardys, E. Boersma, and D. Rizopoulos. Joint models with multiple longitudinal outcomes and a time-to-event outcome: a corrected two-stage approach. *Statistics and Computing*, pages 1–16, 2020. 50

G. McLachlan and T. Krishnan. *The EM Algorithm and Extensions*. Wiley Series in Probability and Statistics. Wiley, 1996. ISBN 9780471123583. URL `https://books.google.gr/books?id=iRSWQgAACAAJ`. 40

J. W. Mellors, L. A. Kingsley, C. R. Rinaldo, J. A. Todd, B. S. Hoo, R. P. Kokka, and P. Gupta. Quantitation of hiv-1 rna in plasma predicts outcome after seroconversion. *Annals of Internal Medicine*, 122(8):573–579, 1995. doi: 10.7326/0003-4819-122-8-199504150-00003. URL https://www.acpjournals.org/doi/abs/10.7326/0003-4819-122-8-199504150-00003. PMID: 7887550. 9

J. W. Mellors, C. R. Rinaldo, P. Gupta, R. M. White, J. A. Todd, and L. A. Kingsley. Prognosis in hiv-1 infection predicted by the quantity of virus in plasma. *Science*, 272(5265):1167–1170, 1996. ISSN 00368075, 10959203. URL http://www.jstor.org/stable/2889946. 9

G. Molenberghs and G. Verbeke. *Linear mixed models for longitudinal data*. Springer, 2000. 38, 44, 45

G. Molenberghs, M. G. Kenward, and E. Lesaffre. The analysis of longitudinal ordinal data with nonrandom drop-out. *Biometrika*, 84(1):33–44, 03 1997. ISSN 0006-3444. doi: 10.1093/biomet/84.1.33. URL https://doi.org/10.1093/biomet/84.1.33. 45

G. Molenberghs, M. G. Kenward, and E. Goetghebeur. Sensitivity analysis for incomplete contingency tables: the slovenian plebiscite case. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 50(1):15–29, 2001. doi: 10.1111/1467-9876.00217. URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/1467-9876.00217. 23

G. Molenberghs, C. Beunckens, C. Sotto, and M. G. Kenward. Every missingness not at random model has a missingness at random counterpart with equal fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(2):371–388, 2008. doi: 10.1111/j.1467-9868.2007.00640.x. URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2007.00640.x. 12, 22, 23, 62, 235

G. Molenberghs, G. Fitzmaurice, M. Kenward, A. Tsiatis, and G. Verbeke. *Handbook of*

Missing Data Methodology. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. Taylor & Francis, 2014. ISBN 9781439854617. 48

W. J. Morokoff and R. E. Caflisch. Quasi-monte carlo integration. *Journal of Computational Physics*, 122(2):218 – 230, 1995. ISSN 0021-9991. doi: https://doi.org/10.1006/jcph.1995.1209. URL `http://www.sciencedirect.com/science/article/pii/S0021999185712090`. 80, 142

S. I. Mozumder, M. Rutherford, and P. Lambert. Direct likelihood inference on the cause-specific cumulative incidence function: A flexible parametric regression modelling approach. *Statistics in Medicine*, 37(1):82–97, 2018. doi: 10.1002/sim.7498. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.7498`. 58

A. Muñoz, M.-C. Wang, S. Bass, J. M. G. Taylor, L. A. Kingsley, J. S. Chmiel, B. F. Polk, and T. M. A. C. S. Group. Acquired immunodeficiency syndrome (AIDS)-free time after human immunodeficiency virus type 1 (HIV-1) seroconversion in homosexual men. *American Journal of Epidemiology*, 130(3):530–539, 09 1989. ISSN 0002-9262. doi: 10.1093/oxfordjournals.aje.a115367. URL `https://doi.org/10.1093/oxfordjournals.aje.a115367`. 7

D. Nakanjako, A. N. Kiragga, B. S. Musick, C. T. Yiannoutsos, K. Wools-Kaloustian, L. Diero, P. Oyaro, E. Lugina, J. C. Ssali, A. Kambugu, and P. Easterbrook. Frequency and impact of suboptimal immune recovery on first-line antiretroviral therapy within the International Epidemiologic Databases to Evaluate AIDS in East Africa. *AIDS*, 30(12):1913–1922, 07 2016. 9, 11

R. Natarajan and R. E. Kass. Reference bayesian methods for generalized linear mixed models. *Journal of the American Statistical Association*, 95(449):227–237, 2000. ISSN 01621459. URL `http://www.jstor.org/stable/2669540`. 141

J. A. Nelder and R. Mead. A Simplex Method for Function Minimization. *The Computer Journal*, 7(4):308–313, 01 1965. ISSN 0010-4620. doi: 10.1093/comjnl/7.4.308. URL `https://doi.org/10.1093/comjnl/7.4.308`. 45

W. K. Newey and D. McFadden. Large sample estimation and hypothesis testing. *Handbook of econometrics*, 4:2111–2245, 1994. 65, 117

E. N. Njagi, G. Molenberghs, M. G. Kenward, G. Verbeke, and D. Rizopoulos. A characterization of missingness at random in a generalized shared-parameter joint modeling framework for longitudinal and time-to-event data, and sensitivity analysis. *Biometrical Journal*, 56(6):1001–1015, 2014. doi: 10.1002/bimj.201300028. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/bimj.201300028`. 23, 47, 48

T. R. O'Brien, W. A. Blattner, D. Waters, M. E. Eyster, M. W. Hilgartner, A. R. Cohen, N. Luban, A. Hatzakis, L. M. Aledort, P. S. Rosenberg, W. J. Miley, B. L. Kroner, and J. J. Goedert. Serum HIV-1 RNA Levels and Time to Development of AIDS in the Multicenter Hemophilia Cohort Study. *JAMA*, 276(2):105–110, 07 1996. ISSN 0098-7484. doi: 10.1001/jama.1996.03540020027025. URL `https://doi.org/10.1001/jama.1996.03540020027025`. 9

M. C. Paik. The generalized estimating equation approach when data are not missing completely at random. *Journal of the American Statistical Association*, 92(440): 1320–1329, 1997. ISSN 01621459. URL `http://www.jstor.org/stable/2965402`. 33

A. Pakman. *tmg: Truncated Multivariate Gaussian Sampling*, 2015. URL `https://CRAN.R-project.org/package=tmg`. R package version 0.3. 175

G. Pantaleo, C. Graziosi, and A. S. Fauci. The immunopathogenesis of human immunodeficiency virus infection. *New England Journal of Medicine*, 328(5):327–335, 1993. doi: 10.1056/NEJM199302043280508. URL `https://doi.org/10.1056/NEJM199302043280508`. PMID: 8093551. 7

N. Pantazis, G. Touloumi, A. S. Walker, and A. G. Babiker. Bivariate modelling of longitudinal measurements of two human immunodeficiency type 1 disease progression markers in the presence of informative drop-outs. *Journal of the*

Royal Statistical Society: Series C (Applied Statistics), 54(2):405–423, 2005. doi: 10.1111/j.1467-9876.2005.00491.x. URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9876.2005.00491.x. 5, 62, 233, 234

N. Pantazis, K. Porter, D. Costagliola, A. De Luca, J. Ghosn, M. Guiguet, A. M. Johnson, A. D. Kelleher, C. Morrison, R. Thiebaut, L. Wittkop, and G. Touloumi. Temporal trends in prognostic markers of hiv-1 virulence and transmissibility: an observational cohort study. *The Lancet HIV*, 1(3):e119 – e126, 2014. ISSN 2352-3018. doi: https://doi.org/10.1016/S2352-3018(14)00002-2. URL http://www.sciencedirect.com/science/article/pii/S2352301814000022. 10

N. Pantazis, G. Touloumi, L. Meyer, A. Olson, D. Costagliola, A. D. Kelleher, et al. The impact of transient combination antiretroviral treatment in early hiv infection on viral suppression and immunologic response in later treatment. *AIDS*, 30(6): 879–888, 2016. 63

N. Pantazis, V. Papastamopoulos, V. Paparizos, S. Metallidis, G. Adamis, A. Antoniadou, M. Psichogiou, M. Chini, H. Sambatakou, N. V. Sipsas, C. Gogos, G. Chrysos, P. Panagopoulos, O. Katsarou, A. Gikas, G. Touloumi, and AMACS. Long-term evolution of cd4+ cell count in patients under combined antiretroviral therapy. *AIDS (London, England)*, 33(10):1645–1655, August 2019a. ISSN 0269-9370. doi: 10.1097/qad.0000000000002248. URL https://doi.org/10.1097/QAD.0000000000002248. 9

N. Pantazis, C. Thomadakis, J. del Amo, D. A. del Arco, F. M. Burns, I. Fakoya, and G. Touloumi. Determining the likely place of hiv acquisition for migrants in europe combining subject-specific information and biomarkers data. *Statistical Methods in Medical Research*, 28(7):1979–1997, 2019b. doi: 10.1177/0962280217746437. URL https://doi.org/10.1177/0962280217746437. PMID: 29233073. 8, 154

H. D. Patterson and R. Thompson. Recovery of inter-block information when block

sizes are unequal. *Biometrika*, 58(3):545–554, 1971. ISSN 00063444. URL http://www.jstor.org/stable/2334389. 31

J. Phair, A. Saah, and L. Jacobson. Acquired immune deficiency syndrome occurring within 5 years of infection with human immunodeficiency virus type–1: The multicenter aids cohort study. *Journal of Acquired Immune Deficiency Syndromes*, 5(5): 490–496, May 1992. ISSN 1525-4135. 7

J. C. Pinheiro and D. M. Bates. Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics*, 4(1):12–35, 1995. ISSN 10618600. URL http://www.jstor.org/stable/1390625. 37, 142

R. L. Prentice. Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika*, 69(2):331–342, 1982. ISSN 00063444. URL http://www.jstor.org/stable/2335407. 50, 233

R. L. Prentice. Correlated binary regression with covariates specific to each binary observation. *Biometrics*, 44(4):1033–1048, 1988. ISSN 0006341X, 15410420. URL http://www.jstor.org/stable/2531733. 33

W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press, New York, NY, USA, 3 edition, 2007. ISBN 0521880688, 9780521880688. 55, 68, 85

C. Proust-Lima, J.-F. Dartigues, and H. Jacqmin-Gadda. Joint modeling of repeated multivariate cognitive measures and competing risks of dementia and death: a latent process and latent class approach. *Statistics in Medicine*, 35(3):382–398, 2016. doi: 10.1002/sim.6731. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.6731. 6, 55

H. Putter, M. Fiocco, and R. B. Geskus. Tutorial in biostatistics: competing risks and multi-state models. *Statistics in Medicine*, 26(11):2389–2430, 2007. doi: 10.1002/

sim.2712. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.2712. 7, 228, 238

A. Quintero and E. Lesaffre. Comparing hierarchical models via the marginalized deviance information criterion. *Statistics in Medicine*, 37(16):2440–2454, 2018. doi: 10.1002/sim.7649. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.7649. 179

E. R. Brown and J. G. Ibrahim. A bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Biometrics*, 59(2):221–228, 2003. doi: 10.1111/1541-0420.00028. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1541-0420.00028. 53, 55

D. Rizopoulos. Jm: An r package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software, Articles*, 35(9):1–33, 2010. ISSN 1548-7660. doi: 10.18637/jss.v035.i09. URL https://www.jstatsoft.org/v035/i09. 53, 54

D. Rizopoulos. Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive gaussian quadrature rule. *Computational Statistics & Data Analysis*, 56(3):491 – 501, 2012a. ISSN 0167-9473. doi: https://doi.org/10.1016/j.csda.2011.09.007. URL http://www.sciencedirect.com/science/article/pii/S0167947311003264. 50, 53, 55

D. Rizopoulos. *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R*. Chapman & Hall/CRC Biostatistics Series. CRC Press, 2012b. ISBN 9781439872871. URL https://books.google.gr/books?id=E__RBQAAQBAJ. 40, 41, 47, 48, 50, 51, 52, 55, 56, 57, 86, 114, 158, 180, 229, 234

D. Rizopoulos. The r package jmbayes for fitting joint models for longitudinal and time-to-event data using mcmc. *Journal of Statistical Software, Articles*, 72(7):1–46, 2016. ISSN 1548-7660. doi: 10.18637/jss.v072.i07. URL https://www.jstatsoft.org/v072/i07. 53, 54

D. Rizopoulos, G. Verbeke, and E. Lesaffre. Fully exponential laplace approximations for the joint modelling of survival and longitudinal data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(3):637–654, 2009. doi: 10.1111/j.1467-9868.2008.00704.x. URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2008.00704.x. 41

J. M. Robins, A. Rotnitzky, and L. P. Zhao. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89(427):846–866, 1994. ISSN 01621459. URL http://www.jstor.org/stable/2290910. 5

J. M. Robins, A. Rotnitzky, and L. P. Zhao. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, 90(429):106–121, 1995. ISSN 01621459. URL http://www.jstor.org/stable/2291134. 5, 33, 35

P. R. Rosenbaum and D. B. Rubin. Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society: Series B (Methodological)*, 45(2):212–218, 1983. doi: 10.1111/j.2517-6161.1983.tb01242.x. URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1983.tb01242.x. 23

P. R. Rosenbaum and D. B. Rubin. The bias due to incomplete matching. *Biometrics*, 41(1):103–116, 1985. ISSN 0006341X, 15410420. URL http://www.jstor.org/stable/2530647. 23

A. Rotnitzky and D. Wypij. A note on the bias of estimators with missing data. *Biometrics*, 50(4):1163–1170, 1994. ISSN 0006341X, 15410420. URL http://www.jstor.org/stable/2533454. 65

D. Rubin. *Multiple Imputation for Nonresponse in Surveys*. Wiley Series in Probability and Statistics. Wiley, 1987. ISBN 9780471087052. URL https://books.google.gr/books?id=0KruAAAAMAAJ. 5, 34

D. B. Rubin. Inference and missing data. *Biometrika*, 63(3):581–592, 1976. ISSN 00063444. URL http://www.jstor.org/stable/2335739. 4, 5, 16, 18, 21, 113, 170

D. Ruppert, M. P. Wand, and R. J. Carroll. *Semiparametric Regression*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2003. doi: 10.1017/CBO9780511755453. 41

C. Saha and M. P. Jones. Asymptotic bias in the linear mixed effects model under non-ignorable missing data mechanisms. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(1):167–182, 2005. doi: 10.1111/j.1467-9868.2005.00494.x. URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2005.00494.x. 85, 117

S. K. Sahu, D. K. Dey, H. Aslanidou, and D. Sinha. A weibull regression model with gamma frailties for multivariate survival data. *Lifetime data analysis*, 3(2):123–137, 1997. 94

D. Salmerón, J. A. Cano, and M. D. Chirlaque. Reducing monte carlo error in the bayesian estimation of risk ratios using log-binomial regression models. *Statistics in Medicine*, 34(19):2755–2767, 2015. doi: 10.1002/sim.6527. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.6527. 160

B. R. Saville and A. H. Herring. Testing random effects in the linear mixed model using approximate bayes factors. *Biometrics*, 65(2):369–376, 2009. ISSN 0006341X, 15410420. URL http://www.jstor.org/stable/25502297. 115

D. O. Scharfstein, A. Rotnitzky, and J. M. Robins. Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association*, 94(448):1096–1120, 1999. ISSN 01621459. URL http://www.jstor.org/stable/2669923. 5, 23, 35

M. D. Schluchter. Methods for the analysis of informatively censored longitudinal data. *Statistics in Medicine*, 11(14-15):1861–1870, 1992. doi: 10.1002/sim.4780111408.

URL https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.4780111408. 5, 6, 12, 49, 62, 233

S. Seaman and A. Copas. Doubly robust generalized estimating equations for longitudinal data. *Statistics in Medicine*, 28(6):937–955, 2009. doi: 10.1002/sim.3520. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.3520. 114

S. Seaman, J. Galati, D. Jackson, and J. Carlin. What is meant by "missing at random"? *Statistical Science*, 28(2):257–268, 2013. ISSN 08834237, 21688745. URL http://www.jstor.org/stable/43288491. 20, 63

S. Self and Y. Pawitan. Modeling a marker of disease progression and onset of disease. In *AIDS epidemiology*, pages 231–255. Springer, 1992. 54

S. G. Self and K.-Y. Liang. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, 82(398):605–610, 1987. ISSN 01621459. URL http://www.jstor.org/stable/2289471. 115

H. Shi, Y. Cheng, and J.-H. Jeong. Constrained parametric model for simultaneous inference of two cumulative incidence functions. *Biometrical Journal*, 55(1):82–96, 2013. doi: 10.1002/bimj.201200011. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/bimj.201200011. 58

X. Song and C. Y. Wang. Semiparametric approaches for joint modeling of longitudinal and survival data with time-varying coefficients. *Biometrics*, 64(2):557–566, 2008. ISSN 0006341X, 15410420. URL http://www.jstor.org/stable/25502092. 55

X. Song, M. Davidian, and A. A. Tsiatis. A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics*, 58(4):742–753, 2002. ISSN 0006341X, 15410420. URL http://www.jstor.org/stable/3068516. 154

D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. Van Der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639, 2002a. doi: 10.1111/1467-9868.00353. URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/1467-9868.00353. 179

D. J. Spiegelhalter, B. N. G., C. B. P., and V. D. L. Angelika. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639, 2002b. doi: 10.1111/1467-9868.00353. URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/1467-9868.00353. 115

J. D. Stamey, D. M. Young, and J. W. Seaman Jr. A bayesian approach to adjust for diagnostic misclassification between two mortality causes in poisson regression. *Statistics in Medicine*, 27(13):2440–2452, 2008. doi: 10.1002/sim.3134. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.3134. 171

D. S. Stein, J. A. Korvick, and S. H. Vermund. CD4+ Lymphocyte Cell Enumeration for Prediction of Clinical Course of Human Immunodeficiency Virus Disease: A Review. *The Journal of Infectious Diseases*, 165(2):352–363, 02 1992. ISSN 0022-1899. doi: 10.1093/infdis/165.2.352. URL https://doi.org/10.1093/infdis/165.2.352. 8

O. Stirrup, A. Copas, A. Phillips, M. Gill, R. Geskus, G. Touloumi, J. Young, H. Bucher, A. Babiker, and C. C. in EuroCoord. Predictors of cd4 cell recovery following initiation of antiretroviral therapy among hiv-1 positive patients with well-estimated dates of seroconversion. *HIV Medicine*, 19(3):184–194, 2018. doi: 10.1111/hiv.12567. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/hiv.12567. 9

O. T. Stirrup, A. G. Babiker, J. R. Carpenter, and A. J. Copas. Fractional brownian motion and multivariate-t models for longitudinal biomedical data, with application to CD4 counts in HIV-positive patients. *Statistics in Medicine*, 35(9):1514–1532, nov

2015. doi: 10.1002/sim.6788. URL http://dx.doi.org/10.1002/sim.6788. 41, 43, 114, 119, 234

J. Stover, R. Glaubius, L. Mofenson, C. Dugdale, M.-A. Davies, G. Patten, and C. Yiannoutsos. Updates to the spectrum/aim model for estimating key hiv indicators at national and subnational levels. *AIDS*, 33 Suppl 3:1, 09 2019. doi: 10.1097/QAD.0000000000002357. 59, 174, 234

M. A. Tanner and W. H. Wong. The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82(398):528–540, 1987. ISSN 01621459. URL http://www.jstor.org/stable/2289457. 171

J. M. Taylor and N. Law. Does the covariance structure matter in longitudinal modelling for the prediction of future cd4 counts? *Statistics in Medicine*, 17(20):2381–2394, 1998. 41, 114, 234

R. A. Thisted. *Elements of statistical computing: Numerical computation*, volume 1. CRC Press, 1988. 87, 164

C. Thomadakis, L. Meligkotsidou, N. Pantazis, and G. Touloumi. Longitudinal and time-to-drop-out joint models can lead to seriously biased estimates when the drop-out mechanism is at random. *Biometrics*, 75(1):58–68, 2019. doi: 10.1111/biom.12986. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/biom.12986. 229, 236

C. Thomadakis, L. Meligkotsidou, N. Pantazis, and G. Touloumi. Misspecifying the covariance structure in a linear mixed model under mar drop-out. *Statistics in Medicine*, 39(23):3027–3041, 2020. doi: 10.1002/sim.8589. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.8589. 237

G. Touloumi and A. Hatzakis. Natural history of hiv-1 infection. *Clinics in dermatology*, 18:389–99, 07 2000. doi: 10.1016/S0738-081X(99)00134-0. 9

G. Touloumi, S. J. Pocock, A. G. Babiker, and J. H. Darbyshire. Estimation and comparison of rates of change in longitudinal studies with informative drop-outs. *Statistics in Medicine*, 18(10):1215–1233, 1999. doi: 10.1002/(SICI)1097-0258(19990530)18:10⟨1215::AID-SIM118⟩3.0.CO;2-6. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291097-0258%2819990530%2918%3A10%3C1215%3A%3AAID-SIM118%3E3.0.CO%3B2-6`. 5, 6, 12, 49, 62, 65, 233

G. Touloumi, N. Pantazis, A. G. Babiker, S. A. Walker, O. Katsarou, A. Karafoulidou, A. Hatzakis, K. Porter, and CASCADE Collaboration. Differences in hiv rna levels before the initiation of antiretroviral therapy among 1864 individuals with known hiv-1 seroconversion dates. *AIDS (London, England)*, 18(12):1697—1705, August 2004. ISSN 0269-9370. doi: 10.1097/01.aids.0000131395.14339.f5. URL `https://doi.org/10.1097/01.aids.0000131395.14339.f5`. 9, 10

G. Touloumi, N. Pantazis, D. Pillay, D. Paraskevis, M.-L. Chaix, H. C. Bucher, C. Kücherer, R. Zangerle, A.-M. B. Kran, K. Porter, on behalf of the CASCADE collaboration in EuroCoord, R. Zangerle, A. D. Kelleher, D. A. Cooper, P. Grey, R. Finlayson, M. Bloch, T. Kelleher, T. Ramacciotti, L. Gelgor, D. Cooper, D. Smith, J. Gill, L. B. Jørgensen, I. Lutsar, G. Chêne, F. Dabis, R. Thiebaut, B. Masquelier, D. Costagliola, M. Guiguet, P. Vanhems, M.-L. Chaix, J. Ghosn, C. Goujard, L. Meyer, F. Boufassa, O. Hamouda, C. Kücherer, B. Bartmeyer, G. Touloumi, N. Pantazis, O. Katsarou, V. Paparizos, P. Gargalianos-Kakolyris, M. Lazanas, G. Rezza, M. Dorrucci, A. d. Monforte, A. De Luca, M. Prins, R. Geskus, J. van der Helm, H. Schuitemaker, M. Sannes, O. Brubakk, A.-M. Bakken Kran, M. Rosinska, R. Muga, J. Tor, P. Garcia de Olalla, J. Cayla, J. del Amo, S. Moreno, S. Monge, J. Del Amo, J. del Romero, S. Pérez-Hoyos, H. C. Bucher, M. Rickenbach, P. Francioli, R. Malyuta, R. Brettle, G. Murphy, C. Sabin, K. Porter, A. Johnson, A. Phillips, A. Babiker, V. Delpech, and D. Pillay. Impact of HIV-1 Subtype on CD4 Count at HIV Seroconversion, Rate of Decline, and Viral Load Set Point in European Serocon-

verter Cohorts. *Clinical Infectious Diseases*, 56(6):888–897, 12 2012. ISSN 1058-4838. doi: 10.1093/cid/cis1000. URL https://doi.org/10.1093/cid/cis1000. 8

A. B. Troxel, D. P. Harrington, and S. R. Lipsitz. Analysis of longitudinal data with non-ignorable non-monotone missing values. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47(3):425–438, 1998a. doi: 10.1111/1467-9876.00119. URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/1467-9876.00119. 47

A. B. Troxel, S. R. Lipsitz, and D. P. Harrington. Marginal models for the analysis of longitudinal measurements with nonignorable non-monotone missing data. *Biometrika*, 85(3):661–672, 1998b. ISSN 00063444. URL http://www.jstor.org/stable/2337394. 47

A. Tsiatis. *Semiparametric Theory and Missing Data*. Springer Series in Statistics. Springer New York, 2007. ISBN 9780387373454. URL https://books.google.gr/books?id=xqZFi2EMB40C. 35, 114

A. A. Tsiatis and M. Davidian. A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika*, 88(2):447–458, 06 2001. ISSN 0006-3444. doi: 10.1093/biomet/88.2.447. URL https://doi.org/10.1093/biomet/88.2.447. 54

A. A. Tsiatis, U. Dafni, V. DeGruttola, K. J. Propert, R. L. Strawderman, and M. Wulfsohn. The relationship of cd4 counts over time to survival in patients with aids: Is cd4 a good surrogate marker? In *AIDS epidemiology*, pages 256–274. Springer, 1992. 8

A. A. Tsiatis, V. DeGruttola, and M. S. Wulfsohn. Modeling the relationship of survival to longitudinal data measured with error. applications to survival and cd4 counts in patients with aids. *Journal of the American Statistical Association*, 90(429):27–37, 1995. ISSN 01621459. URL http://www.jstor.org/stable/2291126. 50, 54

O. Tymejczyk, E. Brazier, K. Wools-Kaloustian, M.-A. Davies, M. Dilorenzo, A. Edmonds, R. Vreeman, C. Bolton, C. Twizere, N. Okoko, S. Phiri, G. Nakigozi, P. Lelo, P. von Groote, A. H. Sohn, and o. b. o. t. I. C. Nash, Denis. Impact of Universal Antiretroviral Treatment Eligibility on Rapid Treatment Initiation Among Young Adolescents with Human Immunodeficiency Virus in Sub-Saharan Africa. *The Journal of Infectious Diseases*, 222(5):755–764, 11 2019. ISSN 0022-1899. doi: 10.1093/infdis/jiz547. URL `https://doi.org/10.1093/infdis/jiz547`. 11

A. van Sighem, F. Nakagawa, D. De Angelis, C. Quinten, D. Bezemer, E. O. de Coul, M. Egger, F. de Wolf, C. Fraser, and A. Phillips. Estimating hiv incidence, time to diagnosis, and the undiagnosed hiv epidemic using routine surveillance data. *Epidemiology (Cambridge, Mass.)*, 26(5):653, 2015. 9

G. Verbeke, G. Molenberghs, H. Thijs, E. Lesaffre, and M. G. Kenward. Sensitivity analysis for nonrandom dropout: A local influence approach. *Biometrics*, 57(1):7–14, 2001. ISSN 0006341X, 15410420. URL `http://www.jstor.org/stable/2676836`. 23

Y. Wang and J. M. G. Taylor. Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *Journal of the American Statistical Association*, 96(455):895–905, 2001. ISSN 01621459. URL `http://www.jstor.org/stable/2670229`. 53

R. W. M. Wedderburn. Quasi-likelihood functions, generalized linear models, and the Gauss—Newton method. *Biometrika*, 61(3):439–447, 12 1974. ISSN 0006-3444. doi: 10.1093/biomet/61.3.439. URL `https://doi.org/10.1093/biomet/61.3.439`. 32

M. West. Generalized linear models: scale parameters, outlier accommodation and prior distributions. *Bayesian Statistics*, 2:531–558, 1985. 36

S. Wilhelm and B. G. Manjunath. *tmvtnorm: Truncated Multivariate Normal and Student t Distribution*, 2015. URL `http://CRAN.R-project.org/package=tmvtnorm`. R package version 1.4-10. 80

P. R. Williamson, R. Kolamunnage-Dona, P. Philipson, and A. G. Marson. Joint modelling of longitudinal and competing risks data. *Statistics in Medicine*, 27(30): 6426–6438, 2008. doi: 10.1002/sim.3451. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.3451. 6, 56, 57

G. Y. Wong and W. M. Mason. The hierarchical logistic regression model for multilevel analysis. *Journal of the American Statistical Association*, 80(391):513–524, 1985. ISSN 01621459. URL http://www.jstor.org/stable/2288464. 36

M. C. Wu and K. Bailey. Analysing changes in the presence of informative right censoring caused by death and withdrawal. *Statistics in Medicine*, 7(1-2):337–346, 1988. doi: 10.1002/sim.4780070134. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.4780070134. 48

M. C. Wu and K. R. Bailey. Estimation and comparison of changes in the presence of informative right censoring: Conditional linear model. *Biometrics*, 45(3):939–955, 1989. ISSN 0006341X, 15410420. URL http://www.jstor.org/stable/2531694. 46, 48

M. C. Wu and R. J. Carroll. Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics*, 44(1): 175–188, 1988. ISSN 0006341X, 15410420. URL http://www.jstor.org/stable/2531905. 5, 48, 49

M. S. Wulfsohn and A. A. Tsiatis. A joint model for survival and longitudinal data measured with error. *Biometrics*, 53(1):330–339, 1997. ISSN 0006341X, 15410420. URL http://www.jstor.org/stable/2533118. 5, 6, 50, 53, 55, 83, 233

W. Ye, X. Lin, and J. M. G. Taylor. Semiparametric modeling of longitudinal measurements and time-to-event data–a two-stage regression calibration approach. *Biometrics*, 64(4):1238–1246, 2008. doi: 10.1111/j.1541-0420.2007.00983.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1541-0420.2007.00983.x. 52, 54

S. L. Zeger and M. R. Karim. Generalized linear models with random effects; a gibbs sampling approach. *Journal of the American Statistical Association*, 86(413):79–86, 1991. ISSN 01621459. URL http://www.jstor.org/stable/2289717. 3

S. L. Zeger and K.-Y. Liang. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, 42(1):121–130, 1986. ISSN 0006341X, 15410420. URL http://www.jstor.org/stable/2531248. 32

S. L. Zeger, K.-Y. Liang, and P. S. Albert. Models for longitudinal data: A generalized estimating equation approach. *Biometrics*, 44(4):1049–1060, 1988. ISSN 0006341X, 15410420. URL http://www.jstor.org/stable/2531734. 3, 32, 37

D. Zeng and J. Cai. Asymptotic results for maximum likelihood estimators in joint analysis of repeated measurements and survival time. *The Annals of Statistics*, 33(5): 2132–2163, 2005. ISSN 00905364. URL http://www.jstor.org/stable/3448637. 53

D. Zeng, G. Yin, and J. G. Ibrahim. Semiparametric transformation models for survival data with a cure fraction. *Journal of the American Statistical Association*, 101(474): 670–684, 2006. ISSN 01621459. URL http://www.jstor.org/stable/27590726. 161

L. P. Zhao and R. L. Prentice. Correlated binary regression using a quadratic exponential model. *Biometrika*, 77(3):642–648, 1990. ISSN 00063444. URL http://www.jstor.org/stable/2337004. 33